MC3D-AD: A Unified Geometry-aware Reconstruction Model for Multi-category 3D Anomaly Detection

Jiayi Cheng[1], **Can Gao**[1,4*], **Jie Zhou**[2,3,4], **Jiajun Wen**[1], **Tao Dai**[1] and **Jinbao Wang**[2,3,4]
[1]College of Computer Science and Software Engineering, Shenzhen University
[2]School of Artificial Intelligence, Shenzhen University
[3]National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University
[4]Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen University
2005gaocan@163.com

Abstract

3D Anomaly Detection (AD) is a promising means of controlling the quality of manufactured products. However, existing methods typically require carefully training a task-specific model for each category independently, leading to high cost, low efficiency, and weak generalization. This study presents a novel unified model for Multi-Category 3D Anomaly Detection (MC3D-AD) that aims to utilize both local and global geometry-aware information to reconstruct normal representations of all categories. First, to learn robust and generalized features of different categories, we propose an adaptive geometry-aware masked attention module that extracts geometry variation information to guide mask attention. Then, we introduce a local geometry-aware encoder reinforced by the improved mask attention to encode group-level feature tokens. Finally, we design a global query decoder that utilizes point cloud position embeddings to improve the decoding process and reconstruction ability. This leads to local and global geometry-aware reconstructed feature tokens for the 3D AD task. MC3D-AD is evaluated on two publicly available Real3D-AD and Anomaly-ShapeNet datasets, and exhibits significant superiority over current state-of-the-art single-category methods, achieving 3.1% and 9.3% improvement in object-level AUROC over Real3D-AD and Anomaly-ShapeNet, respectively. The code is available at https://github.com/iCAN-SZU/MC3D-AD.

1 Introduction

Anomaly Detection (AD) is a critical task for quality control in the manufacturing industry. Early research has concentrated on 2D image data and has achieved significant advancements [Zavrtanik *et al.*, 2021; Zhang *et al.*, 2023; You *et al.*, 2022; Lu *et al.*, 2023]. With the increasing demand for high-precision industrial products, 3D-AD [Bergmann *et al.*, 2022; Liu *et al.*, 2023] has garnered growing attention from researchers, and its objective is to identify and localize anomalous points or regions from 3D point cloud data.

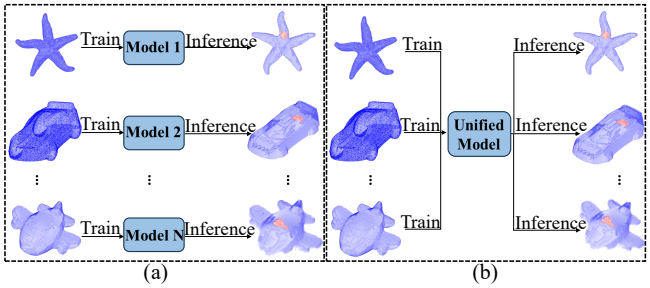

Figure 1: Different settings for 3D anomaly detection. (a) Single-category anomaly detection; (b) Multi-category anomaly detection.

Point cloud data exhibits the characteristics of disorder, sparseness, and structurelessness, which pose great challenges for anomaly detection. Existing 3D-AD methods [Ye *et al.*, 2024] can generally be categorized into: feature embedding-based and reconstruction-based ones. For feature embedding-based approaches, methods like Reg3D-AD [Liu *et al.*, 2023] and Group3-AD [Zhu *et al.*, 2024] have demonstrated their effectiveness by extracting feature embeddings from normal samples. While reconstruction-based methods, such as IMRNet [Li *et al.*, 2024] and R3DAD [Zhou *et al.*, 2024b], focus on learning key features to restore point cloud data, achieving anomaly detection by calculating reconstruction errors.

Additionally, in light of the rich information hidden in multimodal data, multimodal 3D-AD has also attracted much attention from researchers. Some methods, such as BTF [Horwitz and Hoshen, 2023], extract statistical information from both RGB and depth modules to perform anomaly detection. Recently, deep learning-based approaches, such as M3DM [Wang *et al.*, 2023] and CMPF [Cao *et al.*, 2024], have shown promising results by learning feature representations from multimodal data.

Despite achieving very appealing results, some challenges still exist: (1) The trained models are task-specific and lack the generalization to different tasks. In other words, they are required to train an individual and distinct model for each category, seriously limiting their practicality. (2) Existing reconstruction-based methods may fail to learn the in-

*Corresponding author

trinsic features for reconstruction, leading to the problem of "identity shortcut", where the input is directly copied for output without considering its content [You *et al.*, 2022]. As verified in 2D images [You *et al.*, 2022; He *et al.*, 2024; Lu *et al.*, 2023], this phenomenon is significantly amplified in the setting of multi-category anomaly detection. To address these challenges, we propose a unified geometry-aware reconstruction model for 3D anomaly detection named MC3D-AD. Different from previous methods which require training a privately owned model for each category, our method aims to train only one unified model to perform 3D anomaly detection for all categories (See Figure 1).

In light of that point cloud anomalies are usually manifested as irregularities and abnormalities in local geometry structures, we introduce an adaptive geometry-aware masked attention to improve the local feature representation for different categories. It explicitly computes geometric variations within the neighborhood of points and intentionally masks key features selected using their quantified geometric information. To accurately identify anomalies from point cloud data, a transformer-based [Vaswani *et al.*, 2017] architecture is adopted, which incorporates point cloud position embeddings [Li *et al.*, 2023] as global queries and the adaptive geometry-aware masked attention to reconstruct feature tokens for multi-category anomaly detection and localization.

The main contributions of this paper are summarized as follows:

- To perform multi-category 3D anomaly detection, we present a unified framework based on feature reconstruction. To the best of our knowledge, it is the first time to explore multi-category 3D anomaly detection by training only one model.
- To learn robust and generalized representations across categories, we propose a novel adaptive geometry-aware masked attention, which explicitly captures neighborhood geometry information for representation, facilitating the extraction of reconstruction features and also enhancing the interpretability of the model.
- To achieve accurate anomaly detection, we design a local and global geometry-aware transformer, which is reinforced by the proposed adaptive geometry-aware masked attention, thereby providing the ability to reconstruct point clouds from different categories.
- Extensive experiments are conducted to verify the effectiveness of the proposed model, and very impressive results are achieved, with an object-level AUROC improvement of 3.1% over the state-of-the-arts single-category model on Real3D-AD and 9.3% on Anomaly-ShapeNet, respectively.

2 Related Work

2.1 Feature Embedding-Based Methods

Feature embedding-based methods[Liang *et al.*, 2025] extract features from normal samples to form a memory bank using pre-trained models and identify anomalies by comparing test sample features with those in the memory bank. Reg3D-AD [Liu *et al.*, 2023] used the pre-trained PointMAE [Pang *et al.*, 2022] to extract normal features from registered point cloud data and constructed a memory bank to store both global geometric and local coordinate features for anomaly detection. Group3AD [Zhu *et al.*, 2024] introduced contrastive learning for clustering groups to ensure intra-cluster compactness and inter-cluster uniformity, leveraging group-level features stored in a memory bank to detect anomalies. 3D-ST [Bergmann and Sattlegger, 2023] adopted a student-teacher framework to perform feature matching between two networks for 3D anomaly detection. M3DM [Wang *et al.*, 2023] utilized contrastive learning to align RGB and depth modalities, creating a fused representation and a three-level memory bank to jointly enhance detection performance. CPMF [Cao *et al.*, 2024] projected 3D point clouds into 2D images from multi-view and considered image features as global semantic information to complement 3D features, thereby establishing a pseudo multimodal memory bank for anomaly detection. PointAD [Zhou *et al.*, 2024a] aligned local and global features extracted from 2D projections of 3D point clouds using a pre-trained vision-language model, enabling zero-shot 3D anomaly detection.

2.2 Reconstruction-Based Methods

Reconstruction-based methods try to encode normal point cloud data into informative feature representation and restore these features into the original form, with points exhibiting high reconstruction errors identified as anomalies. IMRNet [Li *et al.*, 2024] enhanced PointMAE by incorporating geometric-preserving downsampling and random masking to improve reconstruction fidelity for anomaly detection. R3DAD [Zhou *et al.*, 2024b] leveraged PointNet to iteratively reconstruct fully masked point clouds using a diffusion process, enabling precise localization of abnormal regions. Shape-Guided [Chu *et al.*, 2023] introduced dual memory banks to store normal features extracted from RGB and 3D modalities and reconstructed the input sample at the feature level to achieve robust anomaly detection. Although achieving encouraging results, these methods need to train a task-specific model for each category. Therefore, it is highly desired to develop a unified all-in-one model for all categories.

3 The Proposed Approach

3.1 Problem Description

For multi-category 3D anomaly detection, the available data in the training phase contains point cloud samples from multiple categories, i.e., $P_{train} = \{P^1_{train}, P^2_{train}, \cdots, P^c_{train}\}$, where P^i_{train} denotes the training data from the i-th category and only have normal samples, and c is the number of categories. In the testing phase, the data to be detected includes both normal and anomalous point cloud samples from different categories, i.e., $P_{test} = \{P^1_{test}, P^2_{test}, \cdots, P^c_{test}\}$. The objective is to train a unified model for multiple categories using only normal training data.

3.2 Overview Framework

A key challenge in multi-category 3D anomaly detection is to develop a unified representation method to simultaneously adapt to different categories. To this end, we propose a novel

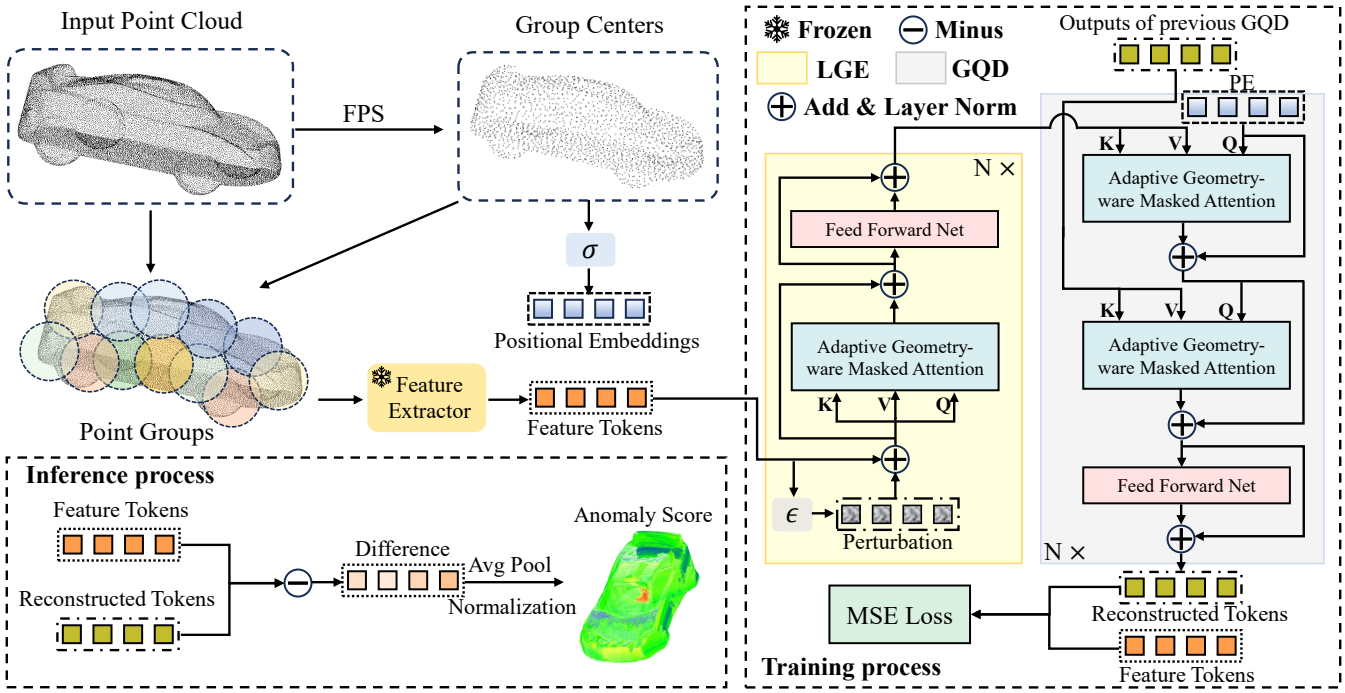

Figure 2: The overview of the proposed method. The input point cloud sample is first registered and aggregated into point groups, while the group centers are projected to obtain position embeddings, and point groups are fed into the feature extractor to generate feature tokens. Then, these tokens and position embeddings are input into the local and global geometry-aware reconstruction framework strengthened by the Adaptive Geometry-aware Masked Attention (AGMA). Finally, anomalies are detected by comparing the differences between the reconstructed and original feature tokens.

geometry-aware reconstruction framework for multi-category 3D anomaly detection. The overall framework is presented in Figure 2, which consists of three main components: Adaptive Geometry-aware Masked Attention (AGMA), Local Geometry-aware Encoder (LGE), and Global Query Decoder (GQD). Each component is described in the following sections.

3.3 Adaptive Geometry-Aware Masked Attention

The representation of normal point cloud data is the key factor for the success of 3D anomaly detection. Existing reconstruction-based methods use the mask attention mechanism to improve the representation ability. Nevertheless, due to the high variation and complexity of point cloud data, they may fail to learn the intrinsic features to represent normal point cloud samples, leading to the phenomenon of "identity shortcut". To address this problem, we propose an AGMA module, which explicitly computes neighborhood geometry information for representation, providing not only better reconstruction ability but also model interpretability.

Specifically, given a point cloud $P = \{p_1, p_2, \cdots, p_n\}$ with each $p_i \in \mathbb{R}^3$, the group centers $\overline{P}_{\text{center}} = \{\overline{p}_1, \overline{p}_2, \cdots, \overline{p}_m\}$ can be obtained by sampling the point cloud through the Furthest Point Sampling (FPS), which can be expressed as:

$$\overline{P}_{\text{center}} = \text{FPS}(P). \tag{1}$$

To capture the local geometric information hidden in neighborhood structure, the adaptive neighborhood of group centers is introduced and can be expressed as:

$$\mathcal{N}_r(\overline{p}_i) = \{\overline{p}_j \in \overline{P}_{\text{center}} \mid \|\overline{p}_i - \overline{p}_j\|_2 \leq r\}, \tag{2}$$

where $\|\cdot\|_2$ denotes the 2-norm, and r represents the neighborhood radius. Due to the varying scales of point clouds across different categories, an adaptive radius is employed to ensure that the number of points in the neighborhood of each group center point remains consistent. The calculation of r can be expressed as:

$$r = \frac{\eta}{|\overline{P}_{\text{center}}|} \sum_{\overline{p}_i \in \overline{P}_{\text{center}}} \|\overline{p}_i - \overline{p}_i^{\text{nearest}}\|_2, \tag{3}$$

where $|\cdot|$ denotes the cardinality of a set, p_i^{nearest} is the nearest neighbor of the p_i, and η is a scaling factor to adjust the size of the neighborhood.

Subsequently, to reflect the geometric information within the neighborhood structure, we define the normal vector and curvature for each group center point $\overline{p}_i$, which can be formulated as:

$$\mathbf{N}_i = FEV_{\min}(\text{Cov}_i), \tag{4}$$

$$C_i = \frac{\lambda_{\min}}{\sum_j^3 \lambda_j}, \tag{5}$$

where $FEV_{\min}$ means finding the eigenvector corresponding to the smallest eigenvalue, λ_i and $\lambda_{\min}$ represents the i-th and the minimum eigenvalue of the covariance matrix, respectively, and Cov_i is the covariance matrix of group center point

$\overline{p}_i$, which is defined as:

$$\text{Cov}_i = \frac{1}{|\mathcal{N}_r(\overline{p}_i)|} \sum_{\overline{p}_i \in \mathcal{N}_r(\overline{p}_i)} (\overline{p}_i - \mu_i)(\overline{p}_i - \mu_i)^T, \quad (6)$$

where μ_i is the centroid of the neighborhood $\mathcal{N}_r(\overline{p}_i)$ of $\overline{p}_i$. Based on the calculated normal vector $\mathbf{N}_i \in \mathbb{R}^3$ and curvature C_i, we can further define an index to quantify the variation in geometric information:

$$Var_i^{\text{geom}} = \alpha Var_i^{\text{norm}} + \beta Var_i^{\text{cruv}}, \quad (7)$$

$$Var_i^{\text{norm}} = \frac{1}{|\mathcal{N}_r(\overline{p}_i)|} \sum_{\overline{p}_j \in \mathcal{N}_r(\overline{p}_i)} \angle(\mathbf{N}_i, \mathbf{N}_j), \quad (8)$$

$$Var_i^{\text{cruv}} = \frac{1}{|\mathcal{N}_r(\overline{p}_i)|} \sum_{\overline{p}_j \in \mathcal{N}_r(\overline{p}_i)} |C_i - C_j|, \quad (9)$$

where $\angle(\cdot,\cdot)$ denotes the angle between two vectors, Var_i^{norm}, Var_i^{cruv}, and Var_i^{geom} represent the degree of change in the normal vector, curvature, and geometric information within the adaptive neighborhood, respectively, and the α and β are hyper-parameters to balance the values.

Intuitively, this geometric change information can be used to guide the learning of representation for reconstruction. Therefore, an improved mask attention mechanism shown in Figure 3 is introduced, aiming to mask some key feature tokens to enhance representation ability.

Specifically, geometric variation information is first extracted for group center points. Then, points with larger and smaller values are randomly selected, respectively, according to the ratio ρ of tokens to be masked, and their corresponding group feature tokens are masked during attention calculation. With this mask attention mechanism, feature representation ability can be significantly improved, and the interpretability of the model is accordingly enhanced.

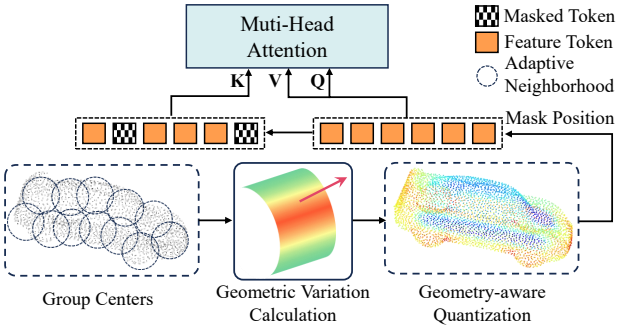

Figure 3: The pipeline of AGMA. The AGMA explicitly extracts geometric variation information from the group center points for mask attention, thereby enhancing the reconstruction representation ability and interpretability.

3.4 Local Geometry-Aware Encoder

Since anomalies usually account for a small area of the entire point cloud, thus encoding local features is essential for improving anomaly detection performance. To encode local group features with geometric information, we propose an LGE to incorporate neighborhood geometric variation information into the feature encoding process. Specifically, a point group is formed by applying k-Nearest Neighbors (KNN) to each group center point $\overline{p}_i$, which can be defined as:

$$G_i = KNN(\overline{p}_i, P). \quad (10)$$

This operation is similar to patch extraction in 2D images, aiming to extract local features from data sequences. Subsequently, the point groups $G = \{G_1, G_2, \cdots, G_g\}$ are input into the feature extractor $\mathcal{F}$ to generate group-level feature tokens $\boldsymbol{F}_{\text{tok}} \in \mathbb{R}^{g \times c}$, where g is the number of groups and c is the number of channels. Concurrently, the group centers P_{center} are passed through a MultiLayer Perceptron (MLP) σ to obtain position embeddings $\boldsymbol{F}_{\text{pos}} \in \mathbb{R}^{g \times c}$, which serve as the Transformer's positional encodings. Additionally, before inputting the feature tokens $\boldsymbol{F}_{\text{tok}}$ into the LGE, feature jittering[You *et al.*, 2022; Bengio *et al.*, 2013] is adopted by adding perturbation to the features. This promotes the representation and reconstruction capability through the denoising process. The perturbation for the i-th group feature can be expressed as:

$$\epsilon^i \sim N\big(0, (\gamma \frac{||\boldsymbol{F}^i{}_{tok}||_2}{C})^2\big), \quad (11)$$

where $N(\cdot,\cdot)$ denotes the normal distribution, and γ is a scaling factor to control the intensity of perturbation on the group feature token.

Finally, the group feature tokens with added noise are fed into the LGE for local feature encoding. The LGE is composed of N sequential blocks, each consisting of an AGMA module and a Feed Forward Network (FNN), where the FNN is implemented by a 2-layer fully connected MLP. By incorporating the AGMA module, the LGE module can encode robust local features with geometric awareness, facilitating subsequent decoding and reconstruction.

3.5 Global Query Decoder

Effectively encoding local features only does not ensure accurate and complete anomaly localization. Moreover, it is evident that global information can positively guide the reconstruction process and improve decoding ability. To this end, we propose a GQD module, which leverages global queries to improve anomaly localization. Specifically, the previously obtained position embeddings $\boldsymbol{F}_{\text{pos}}$ are considered as global queries to be fed into an AGMA. Subsequently, the results are further added with the position embeddings again and fed into another AGMA, followed by an FNN as in the LGE.

To improve the decoding and reconstruction ability, the GQD stacks N repeated blocks. In each block, the local encoding features from the LGE and the position embeddings acted as global queries are input into the first AGMA. Then, the output of the previous block is combined with the local-global features obtained from the first AGMA and fed into the second AGMA. This interaction promotes feature fusion between blocks and facilitates feature decoding and reconstruction. Finally, the GQD outputs the reconstructed feature tokens $\boldsymbol{F}_{\text{rec}} \in \mathbb{R}^{g \times c}$, which is optimized by the MSE loss:

(a) O-AUROC(↑)

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Average
BTF(Raw)	0.520	0.560	0.462	0.432	0.545	0.784	0.549	0.648	0.779	0.754	0.575	0.630	0.603
BTF(FPFH)	0.730	0.647	0.703	0.789	0.707	0.691	0.602	**0.686**	0.596	0.396	0.530	0.539	0.635
M3DM(PointBERT)	0.407	0.506	0.442	0.673	0.627	0.466	0.556	0.617	0.494	0.577	0.528	0.562	0.538
M3DM(PointMAE)	0.434	0.541	0.450	0.683	0.602	0.433	0.540	0.644	0.495	0.694	0.551	0.552	0.552
PatchCore(FPFH)	**0.882**	0.590	0.565	0.837	0.574	0.546	0.675	0.370	0.505	0.589	0.441	0.541	0.593
PatchCore(FPFH+Raw)	0.848	**0.777**	0.626	**0.853**	0.784	0.628	0.837	0.359	0.767	0.663	0.471	0.570	0.682
PatchCore(PointMAE)	0.726	0.498	0.585	0.827	0.783	0.489	0.630	0.374	0.539	0.501	0.519	0.663	0.594
CPMF	0.632	0.518	0.718	0.640	0.640	0.554	0.840	0.349	**0.843**	0.393	0.526	**0.845**	0.625
IMRNet	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	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	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	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
Ours	**0.850**	**0.749**	**0.830**	0.715	**0.955**	**0.831**	0.865	0.560	0.716	**0.803**	**0.766**	0.738	**0.782**

(b) P-AUROC(↑)

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Average
BTF(Raw)	0.564	0.647	0.735	0.608	0.563	0.601	0.514	0.597	0.520	0.489	0.392	0.623	0.571
BTF(FPFH)	**0.738**	0.708	**0.864**	0.693	**0.882**	**0.875**	0.709	**0.891**	0.512	0.571	0.501	0.815	0.730
M3DM(PointBERT)	0.523	0.593	0.682	**0.790**	0.594	0.668	0.589	0.646	0.574	0.732	0.563	0.677	0.636
M3DM(PointMAE)	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(FPFH)	0.471	0.643	0.637	0.618	0.760	0.430	0.464	**0.830**	0.544	0.596	0.522	0.411	0.577
PatchCore(FPFH+Raw)	0.556	0.740	0.749	0.558	0.854	0.658	0.781	0.539	0.808	0.753	0.613	0.549	0.680
PatchCore(PointMAE)	0.579	0.610	0.635	0.683	0.776	0.439	0.714	0.514	0.660	0.725	0.641	0.727	0.642
CPMF	0.618	**0.836**	0.734	0.559	0.753	0.719	**0.988**	0.449	**0.962**	0.725	**0.800**	**0.959**	**0.758**
Reg3D-AD	0.631	0.718	0.724	0.676	0.835	0.503	0.826	0.545	0.817	**0.811**	0.617	0.759	0.705
Group3AD	**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
Ours	0.628	**0.819**	**0.910**	0.640	**0.942**	**0.822**	**0.932**	0.458	0.659	0.778	**0.690**	**0.934**	**0.768**

Table 1: The experimental results for anomaly detection across 12 categories of Real3D-AD. The best and the second-best results are highlighted in **red** and **blue**, respectively. The results of the baselines are excerpted from their papers.

$$\mathcal{L}_{=}\frac{1}{g}\sum_{i=1}^{g}\|\boldsymbol{F}_{\text{tok}}^{i}-\boldsymbol{F}_{\text{rec}}^{i}\|_{2}. \tag{12}$$

During the testing phase, the test point cloud is first registered and grouped for the feature extractor to generate feature tokens. Then, the proposed reconstruction model tries to encode and decode them into the original form. The reconstruction difference is normalized and subjected to average pooling to obtain the final pixel-level anomaly score S_p, which can be expressed as:

$$S_p = AvgPool\big(Norm(\| \boldsymbol{F}_{\text{rec}} - \boldsymbol{F}_{\text{tok}} \|_2)\big), \tag{13}$$

where $Norm$ denotes the min-max normalization, and $AvgPool$ means the operation of average pooling with the kernel size of 1*512. The anomaly score indicates the likelihood of the point being anomalous, and the maximum value of S_p is used as the object-level anomaly score S_o.

4 Experiment

4.1 Experiment Settings

Datasets. (1) Real3D-AD [Liu *et al.*, 2023] is a high-resolution point cloud anomaly detection dataset consisting of 1,254 samples from 12 object categories. Each category has only four training samples but contains anomalies with varying shapes and sizes. (2) Anomaly-ShapeNet [Li *et al.*, 2024] is a synthetic point cloud anomaly detection dataset containing 1,600 samples across 40 categories. Each sample contains between 8,000 and 30,000 points, with the anomalous region accounting for 1% to 10% of the entire point cloud. Due to the large number of categories, this dataset is more challenging for multi-class anomaly detection.

Comparison Methods. Our method adopts the setting of multi-category anomaly detection, where only one model is uniformly trained for all categories. Because existing methods can not apply to multiple categories directly, the compared methods in the experiment use the single-category configuration, wherein privately owned models are separately trained for each category. The proposed method is compared with some representative methods, including BTF [Horwitz and Hoshen, 2023], M3DM [Wang *et al.*, 2023], PatchCore[Roth *et al.*, 2022], CMPF [Cao *et al.*, 2024], Reg3D-AD[Liu *et al.*, 2023], Group3AD [Zhu *et al.*, 2024], IMRNet [Li *et al.*, 2024], and R3D-AD [Zhou *et al.*, 2024b].

Implementation Details. PointMAE pre-trained on ModelNet408K [Wu *et al.*, 2015] is adopted as the feature extractor of our method. The AdamW optimizer is used in the training process, and the learning rate is initially set to 0.0001 and dropped to 0.00001 after 800 epochs. The batch size and the maximum number of epochs are set to 1 and 1000, respectively. The hyperparameters α, β, η, and ρ for AGMA are set to 1, 10, 7, and 0.4, respectively. The number of stacked blocks N in LGE and GQD is set to 4. Our method is performed on PyTorch 1.13.0 and CUDA 11.7 with an NVIDIA A100-PCIE-40GB GPU.

Evaluation Metrics. In the experiments, the Area under the Receiver Operating Characteristic Curve (AUROC, ↑) is used to assess the performance of object-level anomaly detection and pixel-level anomaly localization.

O-AUROC(↑)														
Method	**cap0**	**cap3**	**helmet3**	**cup0**	**bowl4**	**vase3**	**headset1**	**eraser0**	**vase8**	**cap4**	**vase2**	**vase4**	**helmet0**	**bucket1**
BTF(Raw)	0.668	0.527	0.526	0.403	0.664	0.717	0.515	0.525	0.424	0.468	0.410	0.425	0.553	0.321
BTF(FPFH)	0.618	0.522	0.444	0.586	0.609	0.699	0.490	0.719	0.668	0.520	0.546	0.510	0.571	0.633
M3DM	0.557	0.423	0.374	0.539	0.464	0.439	0.617	0.627	0.663	**0.777**	0.737	0.476	0.526	0.501
Patchcore(FPFH)	0.580	0.453	0.404	0.600	0.494	0.449	0.637	0.657	0.662	0.757	0.721	0.506	0.546	0.551
Patchcore(PointMAE)	0.589	0.476	0.424	0.610	0.501	0.460	0.627	0.677	0.663	0.727	0.741	0.516	0.556	0.561
CPMF	0.601	0.551	0.520	0.497	0.683	0.582	0.458	0.689	0.529	0.553	0.582	0.514	0.555	0.601
Reg3D-AD	0.693	0.725	0.367	0.510	0.663	0.650	0.610	0.343	0.620	0.643	0.605	0.500	0.600	0.752
IMRNet	0.737	**0.775**	0.573	0.643	0.676	0.700	0.676	0.548	0.630	0.652	0.614	0.524	0.597	**0.771**
R3D-AD	**0.822**	**0.730**	**0.707**	**0.776**	**0.744**	**0.742**	**0.795**	**0.890**	**0.721**	0.681	**0.752**	**0.630**	**0.757**	0.756
Ours	**0.793**	0.701	**0.979**	**0.743**	**0.911**	**0.761**	**0.886**	**0.776**	**0.670**	**0.835**	**0.929**	**0.876**	**0.672**	**0.784**

Method	**bottle3**	**vase0**	**bottle0**	**tap1**	**bowl0**	**bucket0**	**vase5**	**vase1**	**vase9**	**ashtray0**	**bottle1**	**tap0**	**phone**	**cup1**
BTF(Raw)	0.568	0.531	0.597	0.573	0.564	0.617	0.585	0.549	0.564	0.578	0.510	0.525	0.563	0.521
BTF(FPFH)	0.322	0.342	0.344	0.546	0.509	0.401	0.409	0.219	0.268	0.420	0.546	0.560	0.671	0.610
M3DM	0.541	0.423	0.574	0.739	0.634	0.309	0.317	0.427	0.663	0.577	0.637	**0.754**	0.357	0.556
Patchcore(FPFH)	0.572	0.455	0.604	0.766	0.504	0.469	0.417	0.423	0.660	0.587	0.667	0.753	0.388	0.586
Patchcore(PointMAE)	0.650	0.447	0.513	0.538	0.523	0.593	0.579	0.552	0.629	0.591	0.601	0.458	0.488	0.556
CPMF	0.405	0.451	0.520	0.697	0.783	0.482	0.618	0.345	0.609	0.353	0.482	0.359	0.509	0.499
Reg3D-AD	0.525	0.533	0.486	0.641	0.671	0.610	0.520	0.702	0.594	0.597	0.695	0.676	0.414	0.538
IMRNet	0.640	0.533	0.552	0.696	0.681	0.580	0.676	**0.757**	0.594	0.671	0.700	0.676	0.755	**0.757**
R3D-AD	**0.781**	**0.788**	**0.733**	**0.900**	**0.819**	**0.683**	**0.757**	0.729	**0.718**	**0.833**	**0.737**	0.736	**0.762**	**0.757**
Ours	**0.756**	**0.821**	**0.795**	**0.970**	**0.930**	**0.898**	**0.976**	**0.857**	**0.736**	**0.962**	**0.709**	**0.945**	**0.919**	**0.952**

Method	**vase7**	**helmet2**	**cap5**	**shelf0**	**bowl5**	**bowl3**	**helmet1**	**bowl1**	**headset0**	**bag0**	**bowl2**	**jar**	**Mean**
BTF(Raw)	0.448	0.602	0.373	0.164	0.417	0.385	0.349	0.264	0.378	0.410	0.525	0.420	0.493
BTF(FPFH)	0.518	0.542	0.586	0.609	0.699	0.490	0.719	0.668	0.520	0.546	0.510	0.424	0.528
M3DM	0.657	0.623	0.639	0.564	0.409	0.617	0.427	0.663	0.577	0.537	0.684	0.441	0.552
Patchcore(FPFH)	0.693	0.425	**0.790**	0.494	0.558	0.537	0.484	0.639	0.583	0.571	0.615	0.472	0.568
Patchcore(PointMAE)	0.650	0.447	0.538	0.523	0.593	0.579	0.552	0.629	0.591	0.601	0.458	0.483	0.562
CPMF	0.397	0.462	0.697	0.685	0.685	0.658	0.589	0.639	0.643	0.643	0.625	0.610	0.559
Reg3D-AD	0.462	0.614	0.467	0.688	0.593	0.348	0.381	0.525	0.537	0.706	0.490	0.592	0.572
IMRNet	0.635	**0.641**	0.652	0.603	**0.710**	0.599	0.600	0.702	0.720	0.660	0.685	0.780	0.661
R3D-AD	**0.771**	**0.633**	0.670	**0.696**	0.656	**0.767**	**0.720**	**0.778**	**0.738**	**0.720**	**0.741**	**0.838**	**0.749**
Ours	**0.938**	0.609	**0.761**	**0.841**	**0.754**	**0.885**	**1.000**	**0.978**	**0.862**	**0.805**	**0.719**	**0.971**	**0.842**

Table 2: The object-level AUROC experimental results for anomaly detection across 40 categories of Anomaly-ShapeNet. The best and the second-best results are highlighted in **red** and **blue**, respectively. The results of the baselines are excerpted from their papers, and the pixel-level AUROC experimental results are provided in the supplementary material.

4.2 Main Results

Results on Real3D-AD. The comparison results of MC3D-AD and the existing methods are shown in Table 1. It can be observed that MC3D-AD achieved SOTA performance, with an AUROC of 0.782 at the object level and 0.768 at the pixel level. These results are 3.1% and 1.0% higher than those of the second-best method, which is a task-specific model that carefully tunes for each category.

Results on Anomaly-ShapeNet. The experimental results of MC3D-AD on the Anomaly-ShapeNet dataset are presented in Table 2. Despite the increased complexity of this dataset, which comprises 40 categories, MC3D-AD still achieved SOTA performance. Specifically, the object-level AUROC for anomaly detection reached 0.842, outperforming the second-best single-category method by 9.3%. Additionally, the pixel-level AUROC for anomaly localization attained 0.748, which is 8.0% higher than the second-best single-category approach. These results clearly demonstrate the generalization capabilities of MC3D-AD in multi-class anomaly detection setting.

4.3 Ablation Studies

Table 3 presents the results of the experiments evaluating the effectiveness of the proposed modules. Specifically, LGE refers to a model that incorporates only the LGE module, without the AGMA module. LGE$_{AGMA}$ refers to the application of AGMA within the LGE framework, while GQD$_{AGMA}$ indicates the addition of AGMA to the GQD module. The results show that AGMA plays a crucial role in capturing point cloud geometry features, significantly enhancing its reconstruction capability. This leads to a significant increase in anomaly detection performance, with a 9.4% improvement in O-AUROC. On the other hand, LQD provides additional guidance for the model's reconstruction process, further improving anomaly localization and yielding an 8.3% improvement in P-AUROC. AGMA can be seamlessly integrated into both LGE and GQD, resulting in an overall performance enhancement across all metrics.

Method	**O-AUROC** (↑)	**P-AUROC** (↑)
LGE	0.658	0.650
LGE$_{\text{AGMA}}$	0.752	0.709
LGE+GQD	0.741	0.733
LGE+GQD$_{\text{AGMA}}$	0.756	0.755
LGE$_{\text{AGMA}}$+GQD	0.755	0.760
MC3D-AD	**0.782**	**0.768**

Table 3: Ablation results on Real3D-AD.

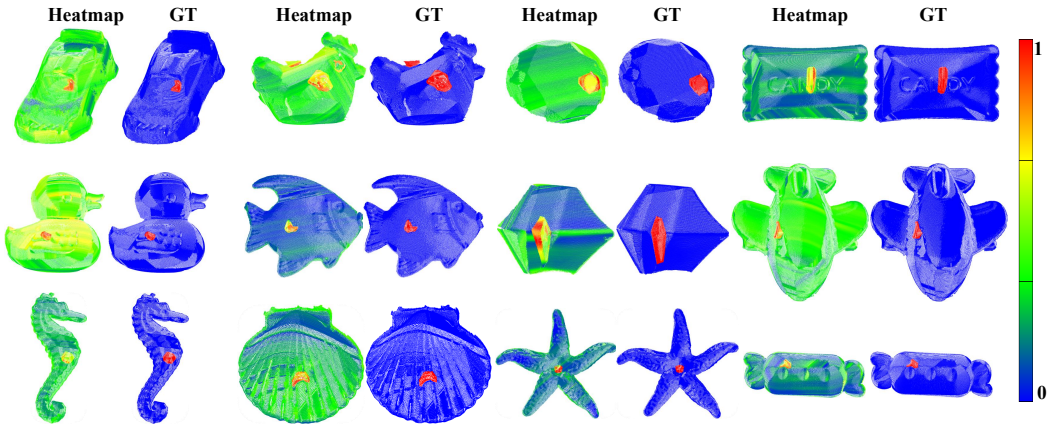

Figure 5: Point heatmap comparison of our MC3D-AD with the Ground Truth (GT) on Real3D-AD. As evidenced by the red-colored areas in the visualized heatmaps, M3DM accurately detects and localizes anomalous regions within the point clouds from different categories.

4.4 Analysis of Hyper-Parameters

Our method introduces two key parameters η and ρ to control the size of the adaptive neighborhood and the proportion of masked tokens, respectively. As shown in Figure 4, an inappropriate η value, whether too small or too large, can lead to inaccurate quantization of geometric variation, so a balanced value $\eta = 7$ is set to ensure stable performance. Similarly, the mask proportion in the attention mechanism requires careful tuning: a low mask ratio hinders the learning of robust reconstruction, while a high mask ratio increases the difficulty of reconstruction. In the experiments, the parameter ρ is set to 0.4, at which the best performance is reached.

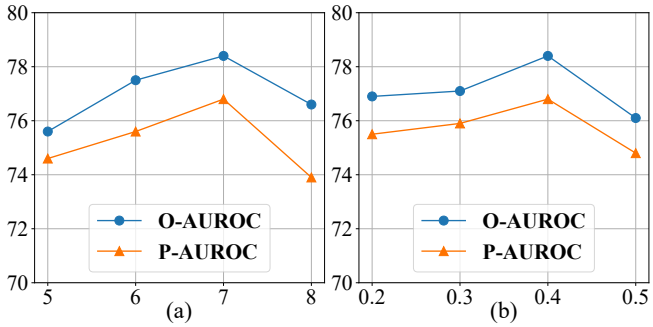

Figure 4: Parameter sensitivity analysis on Real3D-AD. (a) η; (b) ρ.

4.5 Visualization

Figure 5 shows the heatmap visualization results of our MC3D-AD on the Real3D-AD. It is observed that MC3D-AD accurately detects and localizes anomalous regions within the point cloud, clearly demonstrating its effectiveness. In addition, Figure 6 provides the heatmap visualization of the proposed AGMA, which encapsulates the geometric information extracted from the point cloud. The blue regions indicate that the geometric information of the point cloud varies slowly within the adaptive neighborhood, while the transition from green to yellow and red indicates a gradual increase in the change of geometric information within the neighborhood. By intentionally masking blue or red areas during training, the reconstruction ability of our method is greatly improved.

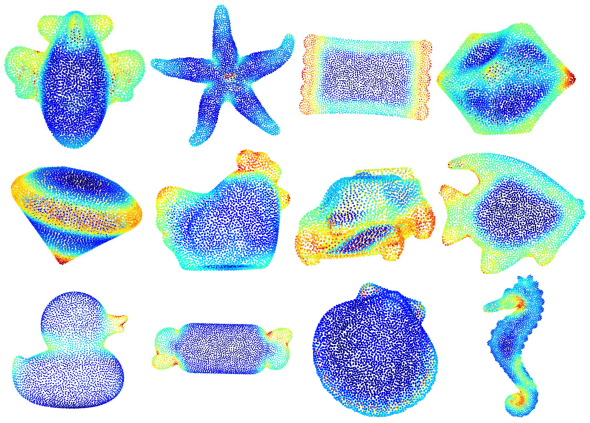

Figure 6: Visualization of our AGMA on Real3D-AD. AGAM extracts geometric variation information from the neighborhood of points, with regions colored in blue indicating gentle changes in geometric structure and areas colored in red exhibiting drastic changes.

5 Conclusion

In this paper, we propose a unified reconstruction framework for multi-category anomaly detection. We introduce an adaptive geometry-aware guided mask attention module, where geometric variation information is captured for robust and generalized representation of different categories. Additionally, we design a geometry-aware transformer with global position embeddings and local mask attention to learn robust reconstructed features. Experiments on benchmark datasets show that our method outperforms existing approaches, achieving state-of-the-art performance. However, further research is needed to exploit the utilization of geometric variation information and develop more robust and efficient frameworks for multi-category anomaly detection.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 62476171, 62476172, and 62206122), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2024A1515011367), the Guangdong Provincial Key Laboratory (Grant No. 2023B1212060076), the Tencent "Rhinoceros Birds" - Scientific Research Foundation for Young Teachers of Shenzhen University, and the Shenzhen Institute of Artificial Intelligence and Robotics for Society.

References

[Bengio *et al.*, 2013] Yoshua Bengio, Li Yao, Guillaume Alain, and Pascal Vincent. Generalized denoising auto-encoders as generative models. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 899–907, 2013.

[Bergmann and Sattlegger, 2023] Paul Bergmann and David Sattlegger. Anomaly detection in 3d point clouds using deep geometric descriptors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2613–2623, 2023.

[Bergmann *et al.*, 2022] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 202–213, 2022.

[Cao *et al.*, 2024] Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024.

[Chu *et al.*, 2023] Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3D anomaly detection. In *Proceedings of the 40th International Conference on Machine Learning*, pages 6185–6194, 2023.

[He *et al.*, 2024] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8472–8480, 2024.

[Horwitz and Hoshen, 2023] Eliahu Horwitz and Yedid Hoshen. Back to the feature: Classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 2968–2977, 2023.

[Li *et al.*, 2023] Zechuan Li, Hongshan Yu, Zhengeng Yang, Tongjia Chen, and Naveed Akhtar. Ashapeformer: Semantics-guided object-level active shape encoding for 3d object detection via transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1012–1021, 2023.

[Li *et al.*, 2024] Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22207–22216, 2024.

[Liang *et al.*, 2025] Hanzhe Liang, Guoyang Xie, Chengbin Hou, Bingshu Wang, Can Gao, and Jinbao Wang. Look inside for more: Internal spatial modality perception for 3d anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5146–5154, 2025.

[Liu *et al.*, 2023] Jiaqi Liu, Guoyang Xie, Ruitao Chen, Xinpeng Li, Jinbao Wang, Yong Liu, Chengjie Wang, and Feng Zheng. Real3d-ad: A dataset of point cloud anomaly detection. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 30402–30415, 2023.

[Lu *et al.*, 2023] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 8487–8500, 2023.

[Pang *et al.*, 2022] Yatian Pang, Wenxiao Wang, Francis E. H. Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *Proceedings of the 16th European Conference Computer Vision*, pages 604–621, 2022.

[Roth *et al.*, 2022] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[Wang *et al.*, 2023] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2023.

[Wu *et al.*, 2015] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015.

[Ye *et al.*, 2024] Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. *arXiv: 2412.12617*, pages 1–13, 2024.

[You *et al.*, 2022] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model

for multi-class anomaly detection. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 4571–4584, 2022.

[Zavrtanik *et al.*, 2021] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. DrÆm – a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, pages 8310–8319, 2021.

[Zhang *et al.*, 2023] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3914–3923, 2023.

[Zhou *et al.*, 2024a] Qihang Zhou, Jiangtao Yan, Shibo He, Wenchao Meng, and Jiming Chen. Pointad: Comprehending 3d anomalies from points and pixels for zero-shot 3d anomaly detection. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 84866–84896, 2024.

[Zhou *et al.*, 2024b] Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *Proceedings of the 18th European Conference Computer Vision*, pages 91–107, 2024.

[Zhu *et al.*, 2024] Hongze Zhu, Guoyang Xie, Chengbin Hou, Tao Dai, Can Gao, Jinbao Wang, and Linlin Shen. Towards high-resolution 3d anomaly detection via group-level feature contrastive learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 4680–4689, 2024.