This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Geometry-Aligned and Anomaly-Aware Reconstruction for 3D Anomaly Detection

Linchun Wu[1] Qin Zou[1,*] Yuanhao Yue[2] Zhongyuan Wang[1]
[1]Wuhan University, China [2]East China Jiaotong University, China
{linchun.wu, qzou, wzy}@whu.edu.cn, yhyue@ecjtu.edu.cn

Abstract

Point cloud anomaly detection is crucial in automated manufacturing, with reconstruction-based diffusion methods emerging as a mainstream solution. However, these approaches still face two major challenges: (1) geometry violation, where random noise perturbations deviate from local surface normals, causing structural distortion; and (2) undistinguished reference regions, where uniformly applied coarse anomaly embeddings during denoising blur normal details and impede accurate anomaly recovery. To address these issues, we propose AARD, a geometry-aligned and anomaly-aware diffusion reconstruction framework. We argue that high-fidelity anomaly detection requires a principled reformulation of the diffusion process: noise should align with geometry to preserve structures, and reconstruction can be better guided by anomaly-aware references to discriminatively recover normal details while correcting defects. AARD progressively aligns noise directions with vertex normals while maintaining vertex-graph consistency, and employs an adaptive transformer that assigns normal references to anomalous regions and input references to normal areas. Experiments on Anomaly-ShapeNet and Real3D-AD show that AARD consistently outperforms state-of-the-art approaches, achieving superior geometric fidelity and robust anomaly localization.

1. Introduction

Point cloud anomaly detection [7, 16] plays a vital role in modern industrial manufacturing [8, 9], safety monitoring [12], and reliable autonomous systems [1]. The objective is to identify and localize subtle deviations from standard geometric patterns, which may correspond to structural defects, potential hazards, or system malfunctions. In these applications, standard patterns establish the normal structural prior of the target object, where normal samples conform to this prior and anomalous ones deviate from it. The core challenge, therefore, is to develop a model that can precisely localize irregular anomalies pattern rather than merely detect acceptable manufacturing differences.

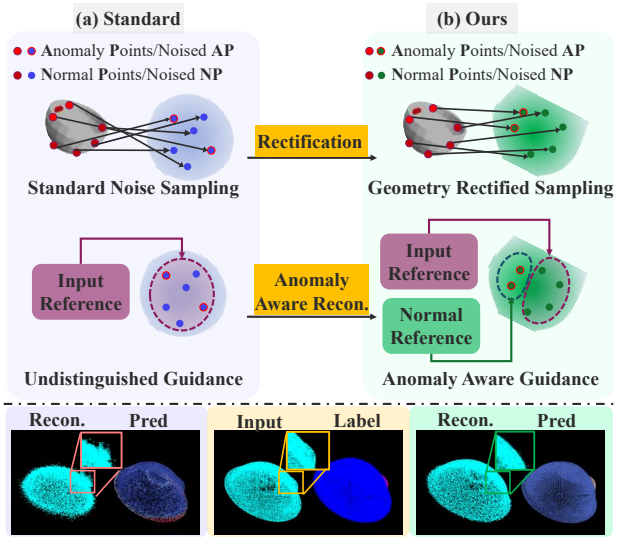

Figure 1. Comparison of reconstruction pipelines. (a) The conventional approach employs standard noise sampling, leading to intersecting noise trajectories and blurred reconstructions. (b) Our proposed method introduces geometry-rectified noise with straightened trajectories and incorporates normal guidance for anomaly-aware reconstruction, achieving higher geometric fidelity.

In this context, reconstruction-based methods have emerged as a prevalent solution. Particularly, the diffusion-based approaches first reconstruct the input point clouds and then detect anomalies by comparing reconstructed samples to their input counterparts. For example, R3D-AD [35] introduces anomaly synthesis for noised anomaly point cloud reconstruction training and employs encoded shape embeddings to enhance geometric fidelity, achieving robust reconstruction quality. Building upon this, MC3AD [5] extend reconstruction to feature level and leverages geometry-normal variation to guide anomaly region early detection, achieving more targeted anomaly restoring and higher

*Corresponding author.

anomaly detection accuracy.

Apart from these recent advancements, two key challenges persist: **(1) Geometric Violation.** In 3D point cloud reconstruction, randomly sampled noise vectors often deviate from local surface normals, producing crossed denosing trajectories. These crossed trajectories yield conflicted reconstruction direction among neighboring vertices—for example, in Figure 1(a), two adjacent anomalous points are perturbed toward opposite directions (left–bottom vs. left–top), ultimately causing fuzzy restoration direction and blurred reconstruction contours, as shown in the bottom row of reconstructed bowl rim. **(2) Undistinguished Reconstruction Guidance.** Existing methods typically apply a uniform coarse input shape embedding to both normal and anomalous regions. However, this unsophisticated strategy presents a critical limitation: normal regions receive insufficient fine-grained structural guidance while anomalous regions lack accurate reconstruction direction. As illustrated in Figure 1(a), the absence of anomaly-aware guidance results in both geometry degradation (blurred contours) and incomplete anomaly correction (residual bulging artifacts).

In this context, we rethink the noise sampling and reconstruction guidance strategies in diffusion-based 3D anomaly detection. Our framework, **AARD** (Geometry-Aligned and Anomaly-Aware Reconstruction), illustrated in Figure 1(b), is driven by two key insights: (1) high-fidelity diffusion reconstruction requires *geometry-aligned noise sampling*, ensuring that noisy trajectories align to local surface structures for a more straightened denoising flow; and (2) effective reconstruction demands *region-specific guidance* rather than uniform priors that indiscriminately blur all areas. To enforce geometry alignment, we introduce a noise rectification module that progressively adjusts random perturbations according to surface-normal directions and vertex-level structural consistency. For region-specific reconstruction, we further design an anomaly aware transformer that dynamically assigns normal-reference guidance to anomalous regions while routing input-reference features to normal regions, thereby preserving fine-grained geometric details. Our main contributions are threefold:

- We present *AARD*, a unified diffusion framework that enhances 3D anomaly reconstruction by integrating geometry-aligned noise sampling with anomaly-aware reference guidance, significantly improving both reconstruction quality and detection accuracy.
- We introduce a sequential noise rectification strategy that ensures straightened denosing flow through geometry-normal alignment and vertex-graph regulation.
- We design an anomaly-aware transformer that disgustingly fuses normal reference features for anomalous regions and input reference features for normal regions, enabling precise anomaly restoration and fine-grained normal preservation.

2. Related Works

2.1. 2D Anomaly Detection

Mainstream 2D anomaly detection approaches can be broadly categorized into reconstruction-based and embedding-based methods. Reconstruction-based methods [17, 22, 26, 29, 32, 34] first learn a reconstruction model exclusively from normal data and then detect anomalies by analyzing the residual differences between the input sample and its reconstructed output. While these methods provide a fundamental solution, their performance often exhibits instability when encountering novel or complex anomaly types that are not represented in the limited (or non-existent) anomaly training set.To address this limitation, synthesis-based methods [4, 10, 13, 28] leverage advancements in image generation frameworks [18, 19, 21, 30] to facilitate anomaly generation, providing additional anomalous training data and thereby improving both reconstruction and detection robustness.

Embedding-based methods utilize pre-trained networks to extract feature representations and compress normal patterns into a compact latent space, where anomaly features become well-separated from normal clusters [2, 25]. Although these methods often achieve strong detection performance, they are typically trained exclusively on normal samples, which limits their generalization ability to unseen anomalies.

2.2. 3D Anomaly Detection

Similarly, current 3D anomaly detection methods can also be broadly categorized into reconstruction-based and feature embedding-based approaches. Reconstruction-based methods [14, 23], primarily based on diffusion models, learn normal patterns mainly from normal data and, in some cases, synthesized anomaly data. Anomalies are then detected by comparing input samples to their reconstructed counterparts. While providing a fundamental solution, these methods often lack explicit normal feature representations, limiting their accuracy in anomaly comparison and detection.

To address this limitation, feature embedding-based methods [16, 23] compare input features against an extensive memory bank of normal samples for robust anomaly detection. Reg3D-AD [16] utilizes memory banks to retrieve the most similar normal samples, while Shape-guided [7] further integrates memory banks from both image and point cloud modalities. These approaches effectively mitigate the distribution gap between training and testing samples, enhancing reconstruction robustness and improving detection interpretability. For multimodal fusion, BTF [9] systematically investigates the role of each modality across various defect types, identifies the optimal feature representations, and strategically combines image

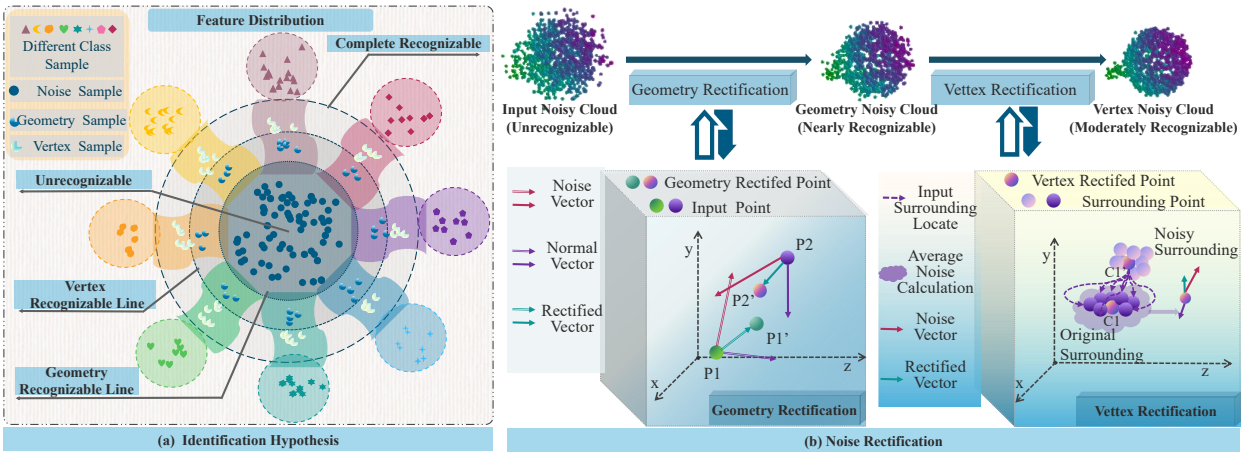

Figure 2. **Illustration of the proposed identification hypothesis (left) and data rectification pipeline (right).** The identification hypothesis partitions the latent space between normal samples and Gaussian noise into four distinct regions: *completely recognizable*, *vertex recognizable*, *geometry recognizable*, and *unrecognizable*. On the right, the data rectification pipeline is depicted. Given a noisy point cloud, *geometry rectification* first aligns noise with local surface normals (e.g., points $P1$ and $P2$), producing geometry-aware rectified points ($P1'$, $P2'$). Subsequently, *vertex rectification* refines each noisy vertex (e.g., $C1$) by averaging the local noise directions of its neighbors, thereby restoring vertex-level structural consistency and yielding refined points ($C1'$).

and point cloud detection, leading to significant improvements in detection accuracy.

While existing diffusion models show promise, they often overlook the critical importance of geometric integrity during noise injection and discriminative capability during denoising. To bridge this gap, we propose AARD, a framework that rectifies the noise direction to adhere to local surface normals and employs an adaptive transformer to provide anomaly-aware references. This dual approach ensures the preservation of fine-grained geometry while enabling precise anomaly recovery.

3. Method

Our method orchestrate a comprehensive denosing pipeline, during noise injection, we design a sequential noise rectification strategy, as shown in Figure 2, to straighten intersected denosing flow; during denosing network processing, we introduce a distinguished anomaly aware transformer, as shown in Figure 3, to achieve precise anomaly region restoration and normal region preservation. Each pipeline detail is illustrated in following sections.

3.1. Noise Rectification

The motivation behind our noise rectification design arises from the observation that 3D meshes exhibit structured geometry with coherent edges and surfaces, whereas the injection of Gaussian noise introduces isotropic perturbations that produce irregular and intersecting denoising trajectories. To regularize this disordered evolution, we establish a theoretical hypothesis that quantifies the point distributions transition from unstructured noise to fully recognizable geometry regarding the geometry recognizability. We denote it as the *Geometry Recognizability Hypothesis*, which guides a straightened denosing trajectories planning with respects to the intrinsic geometry. Specifically, we design a two-stage rectification strategy where geometric rectification aligns noise to vertex normal and vertex-graph rectification ensures local trajectory consistency.

Geometry Recognizability Hypothesis. As illustrated in Figure 2 (a), the core idea of this hypothesis is that, during the diffusion process, the state of a point cloud can evolve continuously from completely unrecognizable random noise to fully recognizable geometry, forming a smoothing transition flow. To begin with, we define the state of a point cloud vertex as a tuple $(\mathbf{p}_s, \mathbf{p}_r)$, where $\mathbf{p}_s \in \mathbb{R}^3$ denotes the spatial position and $\mathbf{p}_r \in \mathbb{R}^3$ represents the surface normal orientation. The *completely unrecognizable state*, corresponding to pure noised point cloud, is defined as:

$$\mathbf{p}_r^{\text{low}} = \mathbf{g}_d, \quad \mathbf{p}_s^{\text{low}} = \mathbf{p}_s + \mathbf{g}_s \cdot \mathbf{p}_r^{\text{low}}, \tag{1}$$

where $\mathbf{g}_s \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I})$ denotes the Gaussian noise magnitude, and $\mathbf{g}_d \sim \mathcal{N}(\mathbf{0}, \sigma_d^2 \mathbf{I})$ represents the directional noise component. Conversely, the *fully recognizable state* with intact geometry is given by:

$$\mathbf{p}_s^{\text{high}} = \mathbf{p}_s, \quad \mathbf{p}_r^{\text{high}} = \mathbf{p}_r. \tag{2}$$

Between these two extremes, we introduce two critical quantitative boundaries to characterize the degree of geometric recognizability:

- **Geometry Recognizable Boundary (GRB):** Defines the minimum structural information required to preserve the

overall geometric integrity. When noise perturbation exceeds this boundary, the global contour and topology become indistinguishable.

- **Vertex Recognizable Boundary (VRB):** Defines an intermediate state that retains local vertex neighborhood relationships while maintaining reconstruction flexibility. This boundary represents the optimal space for detailed restoration.

These boundaries establish quantifiable objectives for noise rectification. Specifically, the GRB boundary ensures global structure preservation and the VRB regularize feature neighboring , which are sequentially achieved with our Geometric Noise Rectification and Vertex-Graph Noise Rectification. Details are illustrated in following sections.

Geometry Recognizable Boundary. To concretize the concept of the *Geometry Recognizable Boundary (GRB)*, we associate it with the discretization of the 3D space. The core idea is that recognizability emerges when the spatial partition is fine enough to distinguish fundamental object parts, yet coarse enough to tolerate moderate noise.

We partition the 3D space into uniform grids of resolution $(n_x, n_y, n_z) \in [-1, 1]$. When the grid is overly coarse, e.g., $g = 2$ with a $2 \times 2 \times 2$ division, an object near the center occupies nearly all cells, leading to invalid over-occupation. Increasing the resolution to $g = 4$ decomposes the object into discernible substructures while preserving its global composition, thereby defining the minimal resolution at which geometry remains perceptually recognizable.

Formally, for each vertex $\mathbf{p}_i \in \mathbb{R}^3$, we introduce isotropic Gaussian noise $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, resulting in an occupancy sphere of radius σ. When a vertex lies near a grid corner, this sphere may extend into adjacent cells. To preserve structural consistency and prevent cross-cell distortion, we constrain the noise such that the occupancy sphere remains within directly neighboring grids. This constraint establishes the normalized noise upper bound:

$$\mathbf{p}_{i,j,k} + \boldsymbol{\sigma}_{\max} < \mathbf{p}_{(i+1),j,k} \quad \Rightarrow \quad \boldsymbol{\sigma}_{\max} < 0.5, \tag{3}$$

where $\boldsymbol{\sigma}_{\max} = (1 - (-1))/g = 0.5$ corresponds to the grid cell size that defines the maximal allowable perturbation preserving fundamental geometric recognizability while allowing denosing trajectory flexibility.

Orthogonal Geometry Rectification. Building upon the established GRB boundary, we introduce a geometry-guided *Orthogonal Rectification* strategy that adaptively regulates both the magnitude and orientation of injected noise with respect to local vertex normals, thereby straightening the denoising trajectory. Specifically, each sampled Gaussian noise vector $\mathbf{g} \in \mathbb{R}^3$ is orthogonally decomposed into two components: one aligned with the vertex normal $\mathbf{n}$ and another lying in the tangent plane:

$$\mathbf{g}_r = \frac{\mathbf{g} \cdot \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} \cdot \mathbf{n}, \quad \mathbf{g}_{\overline{r}} = \mathbf{g} - \mathbf{g}_r, \tag{4}$$

where $\mathbf{g}_r$ denotes the normal-direction perturbation that preserves geometric fidelity, while $\mathbf{g}_{\overline{r}}$ captures tangential deviations that lead to crossed denoising flows and require suppression.

Guided by the *Geometry Recognizable Boundary (GRB)*, we confine the overall perturbation magnitude within $\boldsymbol{\sigma}_{\max} = 0.5$ and emphasize geometry-consistent noise by strengthening $\mathbf{g}_r$ and attenuating $\mathbf{g}_{\overline{r}}$. The rectified vertex state is formulated as:

$$\begin{aligned} \mathbf{p}'_r &= \mathbf{n} \cdot \big(\alpha \mathbf{p}_d + (1-\alpha)\big(\beta \mathbf{g}_r + (1-\beta)\mathbf{g}_{\overline{r}}\big)\big), \\ \mathbf{p}'_s &= \alpha \mathbf{p}_s + (1-\alpha)\mathbf{p}'_r, \end{aligned} \tag{5}$$

where α is a stochastic scaling factor sampled from $\mathcal{U}(0, 0.5)$, and $\beta = \sqrt{\sigma_{\max}^2 - \alpha^2}$ is derived to maintain geometric consistency. Here, $\mathbf{p}_d$ denotes the original magnitude projected along the surface normal.

This rectification process employs the GRB as a hard constraint and reorients random noise under normal guidance, transforming isotropic perturbations into geometry-aligned variations—thereby laying a solid foundation for high-fidelity reconstruction.

Vertex-Graph Noise Rectification. After global geometric rectification, the noise distribution becomes aligned with the object's overall structure. However, local vertex relationships and topological consistency may still deteriorate due to independently sampled perturbations across neighboring points. To address this, we introduce *Vertex-Graph Noise Rectification*, which enforces directional coherence within local vertex neighborhoods, thereby achieving globally normal-aligned and locally conflict-free denoising trajectories.

We define a dynamic **Vertex Group-able Boundary (VGB)**, representing the local structural state where vertices within a group exhibit low directional entropy—indicating that each perturbed point remains spatially coherent with its neighbors, thus forming a recognizable local surface patch.

Specifically, we group vertices into clusters of size $k = 16$ and compute the average noise direction within each cluster:

$$\mathbf{p}_r^{\text{avg}} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{p}_r^i, \tag{6}$$

where $\mathbf{p}_r^i$ denotes the noise direction of vertex i, and $\mathbf{p}_r^{\text{avg}}$ serves as the aggregated directional prior for the group.

The final rectification merges the geometry-guided direction $\mathbf{p}'_r$ with the group-wise average to enforce local regularization:

$$\begin{aligned} \mathbf{p}''_r &= \beta\, \mathbf{p}_r + (1-\beta)\,(\gamma\, \mathbf{p}_r^{\text{avg}} + (1-\gamma)\, \mathbf{p}'_r), \\ \mathbf{p}''_s &= \alpha\, \mathbf{p}_s + (1-\alpha)\, \mathbf{p}''_r, \end{aligned} \tag{7}$$

where $\mathbf{p}''_r$ and $\mathbf{p}''_s$ denote the rectified direction and position, respectively. The weighting parameter $\gamma = 0.5$ is set empirically to balance the influence of the group-wise average and the individual vertex flexibility.

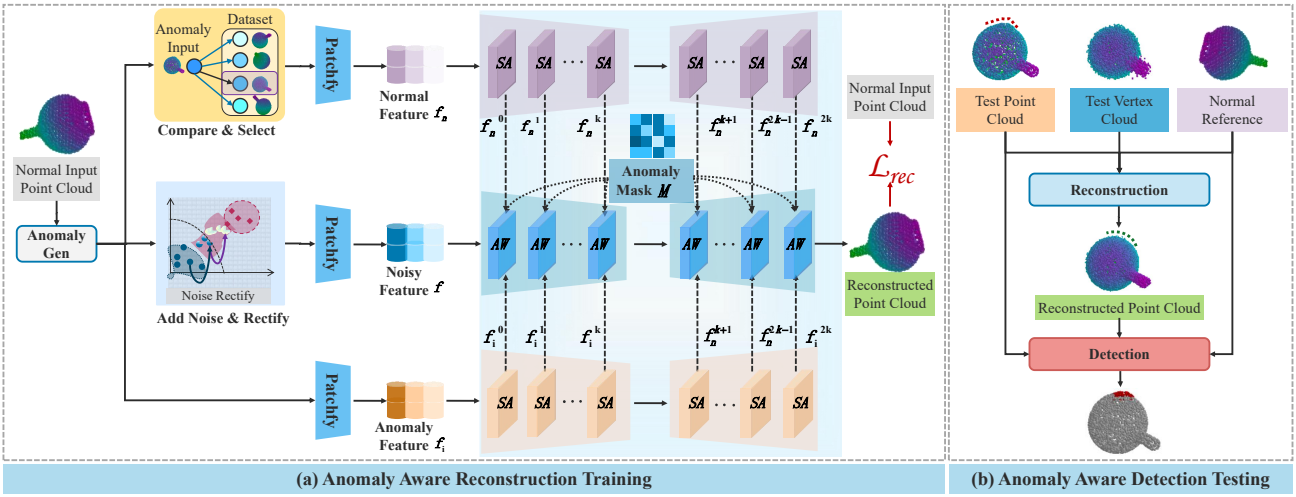

Figure 3. Overview of the proposed AARD framework. During training, normal point clouds are first augmented through anomaly generation. For each input, the most similar normal reference is retrieved, while the corresponding noisy anomaly point cloud is rectified. The rectified point cloud is then processed by the proposed *Anomaly-Aware Transformer*, which integrates input and reference features under the guidance of an adaptive anomaly mask through an anomaly aware attention (AW).

3.2. Anomaly Aware Transformer Network

The aforementioned noise rectification effectively regularizes the denoising trajectory. Beyond this, reconstruction quality can be further enhanced by introducing appropriate reference signals to distinguished spatial region. To this end, we propose an Anomaly-Aware Transformer Network that leverages both normal and anomalous cues for adaptive restoration. As illustrated in Figure 3, the backbone adopts a triplet encoder–decoder architecture, where the input reference and normal reference are processed with respective backbone with self attention (SA), an then fused with noisy features via an *Anomaly-aWare attention mechanism* (AW). Specifically, in the AW, an adaptive anomaly mask is estimated as region partitioning guidance, where normal references is integrated with anomalous features for anomaly restoration, and input references with normal features for fine-grained detail preservation.

For the normal reference selection, we retrieve the most similar sample from the training set using the Chamfer Distance (CD), a widely used metric for evaluating point distribution similarity in 3D generation tasks [24, 31]. To ensure geometric alignment, each candidate is rotated along the x, y, and z axes with a step size of 10° within the range $[-90^\circ, +90^\circ]$, resulting in 5,832 candidates. The sample with the lowest CD relative to the input is then selected as the final normal reference.

Anomaly-Aware Attention. The core objective of this module is to enhance non-defective regions using the input sample—which preserves authentic geometry—while precisely restoring defective areas using the normal reference that contains correct structures. To achieve this, we design a learnable *attention score adaptation* mechanism.

Given the noisy query features $\mathbf{Q} \in \mathbb{R}^{N\times d}$ and the conditional key embeddings $\mathbf{K}_n, \mathbf{K}_i \in \mathbb{R}^{N\times d}$ corresponding to the normal and input references, respectively, we first compute their attention maps:

$$\mathbf{S}_n = \text{Softmax}\left(\mathbf{Q}\mathbf{K}_n^\top\right), \quad \mathbf{S}_i = \text{Softmax}\left(\mathbf{Q}\mathbf{K}_i^\top\right), \tag{8}$$

where $\mathbf{S}_n, \mathbf{S}_i \in \mathbb{R}^{N\times N}$ denote the similarity matrices with respect to the normal and input references.

To localize potential anomalies, we first estimate a coarse anomaly mask $\mathcal{M}_I$ based on the geometric discrepancy between the input and normal reference. We further refine it using attention-level divergence, as $\mathbf{S}_n$ and $\mathbf{S}_i$ share identical queries but differ in keys—thus their discrepancy inherently highlights abnormal regions where attention responses deviate. The adaptive anomaly mask is defined as:

$$\mathcal{M} = \delta(\|\mathbf{P}_i - \mathbf{P}_n\|_2) + \delta(\|\mathbf{S}_i - \mathbf{S}_n\|_2), \tag{9}$$

where $\mathbf{P}_i$ and $\mathbf{P}_n$ denote the input and normal reference point clouds, $\|\cdot\|_2$ is the element-wise Euclidean distance, and $\delta(\cdot)$ is a normalization function (e.g., min–max scaling) that maps values to $[0, 1]$. This adaptive formulation allows the model to refine anomaly localization by jointly leveraging geometric deviation and attention inconsistency. Next, we apply a soft refinement guided by $\mathcal{M}$:

$$\begin{aligned} \hat{\mathbf{S}}_n, \hat{\mathbf{S}}_i &= \text{Softmax}(\text{MLP}([\mathbf{S}_n \,\|\, \mathbf{S}_i])) \odot [\mathcal{M} \,\|\, 1-\mathcal{M}], \\ \tilde{\mathbf{S}}_n &= \hat{\mathbf{S}}_n + \mathbf{S}_n, \quad \tilde{\mathbf{S}}_i = \hat{\mathbf{S}}_i + \mathbf{S}_i, \end{aligned} \tag{10}$$

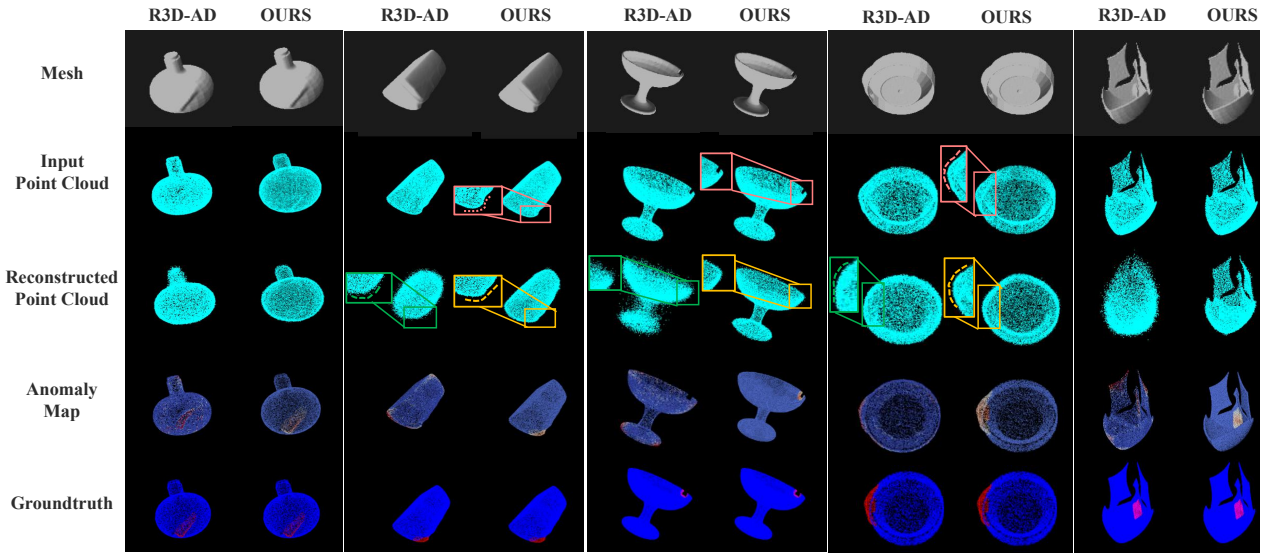

Figure 4. Visual comparison of reconstruction and detection results of R3D-AD and the proposed method on Anomaly-ShapeNet Datasets.

where $[\cdot\|\cdot]$ denotes channel-wise concatenation, and the MLP learns to generate an adaptive attention bias across regions.

Finally, the anomaly-aware attention feature $\mathbf{F}$ is obtained as a weighted fusion of both reference features:

$$\mathbf{F} = \text{Softmax}\left(\frac{\tilde{\mathbf{S}}_n}{\sqrt{d}}\right)\mathbf{V}_n + \text{Softmax}\left(\frac{\tilde{\mathbf{S}}_i}{\sqrt{d}}\right)\mathbf{V}_i, \quad (11)$$

where $\mathbf{V}_n$ and $\mathbf{V}_i$ are the value embeddings derived from the normal and input references, respectively. This dual-reference attention effectively guides the network to restore structural integrity in anomalous regions while maintaining fine details in non-defective areas.

Reconstruction Objective. During training, anomaly samples $\mathbf{P}_i$ are synthetically generated by introducing geometric deviations as $\mathbf{P}_i = \mathbf{P}_n + d$, where $\mathbf{P}_n$ denotes normal input and d represents perturbations, the injected noise is rectified using our proposed geometry and vertex noise sampling strategies as $\epsilon' = r(\epsilon)$, then the overall anomalous noise will be $\mu = \epsilon' + d$. Under this configuration, the reconstruction training objective is formulated as:

$$\mathcal{L}_{rec} = \mathbb{E}_{\varepsilon(\mathbf{P}_n),\, r,\, \mathbf{P}_a,\, \epsilon\sim\mathcal{N}(0,\mathbf{I}),\, t} \left[\|\mu_t - \mu_\theta(\mathbf{z}_t, t, \tau_{\theta_0}(\mathbf{P}_n), \tau_{\theta_1}(\mathbf{P}_i))\|^2 \right], \quad (12)$$

where z_t is the noisy anomaly point cloud at timestep t, and $\tau_\theta(\mathbf{P}_i)$ and $\tau_\theta(\mathbf{P}_n)$ are the encoded anomaly test and normal reference features, respectively. The term r represents the rectification strategy applied to the noise, and $\epsilon_\theta(\cdot)$ is the noise prediction network. This loss encourages the model to accurately predict anomalous noise, leading to improved reconstruction quality and anomaly reconstruction.

Detection Formulation. For anomaly detection, the reconstructed point cloud $\mathbf{P}_r$, the input point cloud $\mathbf{P}$, and the normal reference point cloud $\mathbf{P}_n$ are compared to estimate a normalized geometric discrepancy. The discrepancy is subsequently aggregated through an average pooling operation to produce the final anomaly score, formulated as:

$$S_p = \text{AvgPool}(\text{Norm}(\|\mathbf{P}_r - \mathbf{P}\|_2 + \|\mathbf{P}_r - \mathbf{P}_n\|_2)), \quad (13)$$

where $\mathbf{P}_r$, $\mathbf{P}$, and $\mathbf{P}_n$ denote the reconstructed, test, and reference point clouds, respectively. This formulation jointly measures reconstruction deviation from both the test and normal references, and the subsequent pooling operation effectively aggregates localized geometric inconsistencies into a compact anomaly response map, enabling robust point-level anomaly localization.

4. Experiments

4.1. Datasets

Anomaly-ShapeNet. Anomaly-ShapeNet [14] contains a total of 1,600 samples spanning 40 distinct categories. The dataset includes six types of anomalies: bulge, concavity, crack, holes, and broken. Each category's training set comprises only four samples while the corresponding test sets include both normal and various defective samples.

Real3D-AD. Real3D-AD [16] consists of 1,254 samples distributed across 12 distinct categories, including *Airplane*, *Car*, *Candybar*, *Chicken*, *Diamond*, *Duck*, *Fish*, *Gemstone*, *Seahorse*, *Shell*, *Starfish*, and *Toffees*. Similar to Anomaly-ShapeNet, each category's training set contains only four samples, while the test set comprises normal and various defective samples.

Table 1. The O-AUROC performance (↑) of different methods on Real3D-AD across 12 categories, where best and second-place results are highlighted in **red** and **blue**, respectively.

O-AUROC														
Method	**Airplane**	**Car**	**Candy**	**Chicken**	**Diamond**	**Duck**	**Fish**	**Gemstone**	**Seahorse**	**Shell**	**Starfish**	**Toffees**	**Average**	**Mean Rank**
M3DM (CVPR23')	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	8.417
Reg3D-AD (NeurIPS23')	0.716	0.697	0.685	**0.852**	0.900	0.584	**0.915**	0.417	0.762	0.583	0.506	0.827	0.704	5.833
IMRNet (CVPR24')	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	5.500
R3D-AD (ECCV24')	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	5.542
ISMP (AAAI25')	0.858	**0.731**	**0.852**	0.714	**0.948**	0.712	**0.945**	0.468	0.729	0.623	0.660	0.842	0.757	4.292
PO3AD (CVPR25')	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	5.417
MC4AD	**0.871**	0.684	0.811	0.701	0.836	0.824	0.890	**0.701**	0.754	0.803	**0.770**	0.785	0.786	4.167
PASDF (ICCV25')	0.628	**0.959**	0.788	0.739	0.894	0.658	**0.989**	0.634	**1.000**	**0.850**	0.617	**0.866**	**0.802**	**3.833**
Ours	**0.882**	0.725	**0.864**	**0.871**	**0.917**	**0.846**	0.891	**0.765**	**0.807**	**0.811**	**0.759**	**0.855**	**0.820**	**1.789**

Table 2. P-AUROC performance (↑) of different methods on Real3D-AD across 12 categories, where best and second-place results are highlighted in **red** and **blue**, respectively.

P-AUROC														
Method	**Airplane**	**Car**	**Candy**	**Chicken**	**Diamond**	**Duck**	**Fish**	**Gemstone**	**Seahorse**	**Shell**	**Starfish**	**Toffees**	**Average**	**Mean Rank**
BTF(Raw) (CVPR23')	0.564	0.647	0.735	0.609	0.563	0.601	0.514	0.597	0.520	0.489	0.392	0.623	0.571	8.417
BTF(FPFH) (CVPR23')	0.738	0.708	**0.864**	0.735	0.882	0.875	0.709	0.891	0.512	0.571	0.501	**0.815**	0.733	5.250
M3DM (CVPR23')	0.547	0.602	0.679	0.678	0.608	0.667	0.606	0.674	0.560	0.738	0.532	0.682	0.631	7.417
PatchCore(FPFH) (CVPR22')	0.562	0.754	0.780	0.429	0.828	0.264	0.829	**0.910**	0.739	0.739	0.606	0.747	0.682	6.000
PatchCore(PointMAE) (CVPR22')	0.569	0.609	0.627	0.729	0.718	0.528	0.717	0.444	0.633	0.709	0.580	0.580	0.620	7.417
Reg3D-AD (NeurIPS23')	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	5.667
R3D-AD (ECCV24')	0.594	0.557	0.593	0.620	0.555	0.635	0.573	0.668	0.562	0.578	0.608	0.568	0.592	8.000
ISMP (AAAI25')	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	2.583
MC4AD	**0.792**	**0.839**	0.842	**0.850**	0.844	0.780	**0.934**	0.892	**0.820**	**0.869**	**0.792**	0.784	**0.837**	**2.417**
Ours	**0.835**	0.801	0.857	**0.906**	**0.886**	**0.894**	**0.917**	**0.926**	**0.852**	0.846	**0.785**	0.806	**0.860**	**1.833**

Table 3. O-AUROC performance (↑) of different methods on Anomaly-ShapeNet across 15 categories, where best and second-place results are highlighted in **red** and **blue**, respectively.

O-AUROC																
Method	**ashtray0**	**bag0**	**bottle**	**bowl**	**bucket**	**cap**	**cup**	**eraser0**	**headset**	**helmet**	**jar0**	**micro.0**	**shelf0**	**tap**	**vase**	**Average**
CPMF (PR24')	0.353	0.643	0.498	0.679	0.542	0.601	0.498	0.689	0.551	0.532	0.610	0.509	0.685	0.528	0.514	0.562
Reg3D-AD (NeurIPS23')	0.597	0.706	0.652	0.548	0.681	0.632	0.524	0.343	0.574	0.491	0.592	0.414	0.688	0.659	0.576	0.578
IMRNet (CVPR24')	0.671	0.660	0.666	0.676	0.676	0.704	0.700	0.548	0.698	0.603	0.780	0.755	0.603	0.686	0.629	0.670
R3D-AD (ECCV24')	**0.833**	0.720	0.777	0.751	0.720	0.726	0.767	0.890	0.767	0.704	0.838	0.762	0.696	**0.818**	0.734	0.767
PO3AD (CVPR'25)	**1.000**	0.833	0.917	0.883	0.820	0.800	0.852	0.995	0.866	0.837	0.866	0.776	0.573	0.713	0.826	0.837
Simple3D (AAAI 26')	0.995	0.881	0.975	0.841	0.875	0.858	0.912	**1.000**	**0.969**	0.704	0.905	**1.00**	**0.748**	0.649	0.862	0.860
MC4AD	**1.000**	**0.962**	**0.981**	**0.943**	**0.923**	**0.862**	**0.964**	**1.000**	0.937	**0.898**	**0.914**	0.924	0.696	0.787	**0.889**	**0.912**
Ours	**1.000**	**0.976**	**0.988**	**0.952**	**0.934**	**0.893**	**0.979**	**0.981**	**0.945**	**0.897**	**0.962**	**0.937**	**0.754**	**0.857**	**0.913**	**0.931**

4.2. Implementation Details

The input point clouds are uniformly sampled with 16,384 points, where a k-nearest neighbors (k-NN) clustering is applied with $k = [16, 16]$. The proposed Gradient Transformer adopts a patch size of 2 and a hidden dimension of 1024. The encoder is composed of 16 transformer blocks while the decoder includes 6 transformer blocks. The model is trained using the Adam optimizer with a learning rate of 1×10^{-5} and a batch size of 32. All experiments are conducted on two NVIDIA RTX 3090 GPUs.

4.3. Comparison Against State-of-the-Art Methods

Baselines. To validate the effectiveness of our model, we evaluate our AARD against several state-of-the-art anomaly detection methods, including Simple3D [6] (AAAI 26'), PO3AD [27] (CVPR 25'), PASDF [33] (ICCV 25'), ISMP [15] (AAAI 25'), R3D-AD [35] (ECCV 24'), IMRNet [14] (CVPR 24'), M3DM [23] (CVPR 23'), PatchCore [20] (CVPR 22'), CPMF [3] (PR 24') and RegAD [11] (ECCV 24').

Evaluation Metrics. Following the protocol in [15], we employ the O-AUROC metric for image-level anomaly detection, where higher values indicate better performance. For pixel-level anomaly segmentation, we adopt the pixel-level AUROC (P-AUROC) as the evaluation metric.

Quantitative Results. The evaluations on the Real3D-AD dataset, in terms of O-AUROC and P-AUROC are shown in Table 1 and Table 2,respectively. The proposed model achieving the highest object-level AUROC of 0.82 and pixel-level AUROC of 0.86 . Further comparison results on the Anomaly-ShapeNet dataset, measured by O-AUROC and P-AUROC, are presented in Table 3 and Appendix. Our proposed model achieves the best detection performance,

attaining the highest O-AUROC score of 0.93. These results highlight the significant contributions of our noise rectification strategy and anomaly aware reconstruction network to downstream anomaly inspection tasks.

Qualitative Results. Figure 4 presents a qualitative comparison with the state-of-the-art R3D-AD. Our method reconstructs normal regions with clear geometry and structural integrity, while R3D-AD often yields blurred boundaries and texture distortions. In anomaly detection, R3D-AD produces more false positives, whereas our approach precisely localizes the true anomalous regions.

4.4. Ablation Studies

Impact of Different Components. Ablation studies were conducted on the Anomaly ShapeNet dataset to evaluate the contribution of each component of proposed AARD. The baseline (a) uses a pure Transformer with linear diffusion. We incrementally added Geometry Noise Rectification (GNR), Vertex Noise Rectification (VNR), Normal Reference (NR), Input Reference (IR), and Anomaly Mask (AM) to assess their individual and combined effects.

We apply the O-AUROC metric for anomaly detection, and F-Score for point cloud reconstruction, which is a widely adapted as similarity measurement [24, 31]. As shown in Table 4 (left), the baseline achieves an O-AUROC of 0.77. Incorporating GNR (b) and VNR (c) leads to performance gains of 3% and 2.4%, respectively. Since anomaly mask is indispensable for anomaly aware attention, the incoperation of AM and NR leads to 9% improvemnt while 6.8% for AM and IR. Altogether, the framework reach it's optimized performance of 0.931.

For reconstruction, GNR and VNR improve the F-Score by 11.3%, with NR, IR, and AM adding another 12.1%. As illustrated in Figure 5 (top (c)), for the headset category with highly curved surfaces, the combined rectification effectively straightens the denoising trajectories while remaining adaptive to local curvature variations, producing smooth and authentic geometry. Meanwhile, the normal reference corrects the anomalous bump along the rim, and the input reference preserves the fine-grained point arrangement on the curved surface.

Hyperparameter Sensitivity. We conducted ablation studies on the Anomaly ShapeNet dataset to assess the impact of hyperparameters grid num g of geometry rectification and the group size k of vertex rectification. As shown in Table 4 (bottom), deviating from their optimal values degrades performance: lower values increase noise and hinder reconstruction, while higher values reducing reconstruction flexibility, resulting in suboptimal outcomes.

5. Conclusion

In this paper, we propose a novel Geometry-Aligned and Anomaly-Aware Reconstruction for 3D Anomaly Detection (AARD). For denosing flow straightening, we introduce a sequential rectification mechanisms that not only preserves global geometry but also restores local vertex-to-vertex consistency. Additionally, we design a *Anomaly Aware Transformer* that achieves distinguished normal region preservation and anomaly region reconstruction. Experiments highlights the importance of robust noise rectification and smoother denoising flows for advancing point cloud generation and reconstruction.

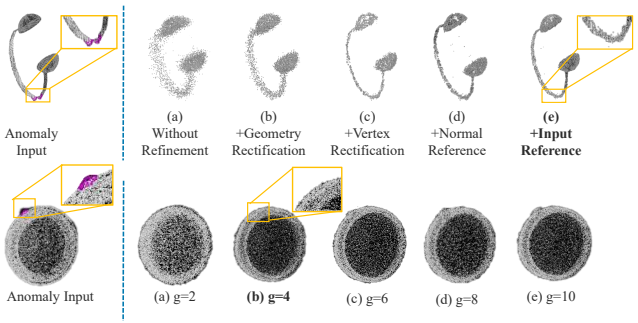

Figure 5. Ablation study on model components (top) and hyperparameter grid num g. For components, the leftmost column shows the input point cloud containing anomalies. Panels (a)–(e) progressively incorporate the proposed components to improve reconstruction performance. For g, when set as 4 achieves best anomaly restoration as well as normal preservation.

Table 4. Ablation studies: Component analysis (top) and hyperparameter sensitivity (bottom) analysis.

Component Analysis (Anomaly ShapeNet).

No.	GNR	VNR	NR	IR	AM	F-Score	O-AUROC
a	×	×	×	×	×	0.719	0.772
b	✓	×	×	×	×	0.815	0.802
c	×	✓	×	×	×	0.791	0.796
d	✓	✓	×	×	×	0.832	0.825
e	✓	✓	✓	×	✓	0.926	0.915
f	✓	✓	×	✓	✓	0.947	0.893
g	✓	✓	✓	✓	✓	**0.953**	**0.931**

Hyperparameter Sensitivity.

Param	**Parameter Values**				
g	2	4	6	8	10
O-AUROC	0.857	**0.931**	0.927	0.915	0.898
k	8	16	24	32	40
O-AUROC	0.891	**0.931**	0.922	0.909	0.886
γ	0.1	0.3	0.5	0.7	0.9
O-AUROC	0.889	0.920	**0.931**	0.927	0.908

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant 62371350 and 62171324.

References

[1] Sihem Baccari, Mohamed Hadded, Hakim Ghazzai, Haifa Touati, and Mourad Elhadef. Anomaly detection in connected and autonomous vehicles: A survey, analysis, and research challenges. *IEEE Access*, 12:19250–19276, 2024. 1

[2] Jaehyeok Bae, Jae-Han Lee, and Seyun Kim. Pni: industrial anomaly detection using position and neighborhood information. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6373–6383, 2023. 2

[3] Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024. 7

[4] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *European Conference on Computer Vision*, pages 37–54. Springer, 2024. 2

[5] Jiayi Cheng, Can Gao, Jie Zhou, Jiajun Wen, Tao Dai, and Jinbao Wang. Mc3d-ad: A unified geometry-aware reconstruction model for multi-category 3d anomaly detection. *arXiv preprint arXiv:2505.01969*, 2025. 1

[6] Yuqi Cheng, Yihan Sun, Hui Zhang, Weiming Shen, and Yunkang Cao. Towards high-resolution 3d anomaly detection: A scalable dataset and real-time framework for subtle industrial defects. *arXiv preprint arXiv:2507.07435*, 2025. 7

[7] Yu-Min Chu, Liu Chieh, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. 2023. 1, 2

[8] Zhihao Gu, Liang Liu, Xu Chen, Ran Yi, Jiangning Zhang, Yabiao Wang, Chengjie Wang, Annan Shu, Guannan Jiang, and Lizhuang Ma. Remembering normality: Memory-guided knowledge distillation for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16401–16409, 2023. 1

[9] Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2968–2977, 2023. 1, 2

[10] Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8526–8534, 2024. 2

[11] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *European Conference on Computer Vision*, pages 303–319. Springer, 2022. 7

[12] Takashi Imabuchi, Yuta Tanifuji, and Kuniaki Kawabata. Discrimination of the structures in nuclear facility by deep learning based on 3d point cloud data. In *2022 IEEE/SICE International Symposium on System Integration (SII)*, pages 1036–1040, 2022. 1

[13] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *IEEE/CVF conference on computer vision and pattern recognition*, pages 9664–9674, 2021. 2

[14] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22207–22216, 2024. 2, 6, 7

[15] 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*, volume 39, pages 5146–5154, 2025. 7

[16] 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. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2, 6

[17] Tongkun Liu, Bing Li, Zhuo Zhao, Xiao Du, Bingke Jiang, and Leqi Geng. Reconstruction from edge image combined with color and gradient difference for industrial surface anomaly detection. *arXiv preprint arXiv:2210.14485*, 2022. 2

[18] Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob Mcgrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. In *International Conference on Machine Learning*, pages 16784–16804, 2022. 2

[19] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 2

[20] 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. 7

[21] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022. 2

[22] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019. 2

[23] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2023. 2, 7

[24] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. *arXiv preprint arXiv:2412.01506*, 2024. 5, 8

[25] Peng Xing, Dong Zhang, Jinhui Tang, and Zechao Li. A recover-then-discriminate framework for robust anomaly de-

tection. *Science China Information Sciences*, 68(4):142102, 2025. 2

[26] Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *European Conference on Computer Vision*, pages 1–17. Springer, 2024. 2

[27] Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1353–1362, 2025. 7

[28] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 2

[29] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021. 2

[30] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *IEEE International Conference on Computer Vision*, pages 5907–5915, 2017. 2

[31] Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiu, Anqi Pang, Haoran Jiang, Wei Yang, Lan Xu, and Jingyi Yu. Clay: A controllable large-scale generative model for creating high-quality 3d assets. *ACM Transactions on Graphics (TOG)*, 43(4):1–20, 2024. 5, 8

[32] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16699–16708, 2024. 2

[33] Bin Zheng, Jialin Gan, Xiaohao Xu, et al. Bridging 3D anomaly localization and repair via high-quality continuous geometric representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 27063–27072, 2025. 7

[34] Yibo Zhou. Rethinking reconstruction autoencoder-based out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7379–7387, 2022. 2

[35] Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *European Conference on Computer Vision*, pages 91–107. Springer, 2025. 1, 7