R3D-AD: Reconstruction via Diffusion for 3D Anomaly Detection

Zheyuan Zhou[1*], Le Wang[1*], Naiyu Fang[1,2], Zili Wang[1], Lemiao Qiu[1(✉)], and Shuyou Zhang[1]

[1] State Key Laboratory of Fluid Power & Mechatronic System, Zhejiang University, Hangzhou, China
[2] S-Lab, Nanyang Technological University, Singapore, Singapore

https://zhouzheyuan.github.io/r3d-ad

Abstract. 3D anomaly detection plays a crucial role in monitoring parts for localized inherent defects in precision manufacturing. Embedding-based and reconstruction-based approaches are among the most popular and successful methods. However, there are two major challenges to the practical application of the current approaches: 1) the embedded models suffer the prohibitive computational and storage due to the memory bank structure; 2) the reconstructive models based on the MAE mechanism fail to detect anomalies in the unmasked regions. In this paper, we propose R3D-AD, reconstructing anomalous point clouds by diffusion model for precise 3D anomaly detection. Our approach capitalizes on the data distribution conversion of the diffusion process to entirely obscure the input's anomalous geometry. It step-wisely learns a strict point-level displacement behavior, which methodically corrects the aberrant points. To increase the generalization of the model, we further present a novel 3D anomaly simulation strategy named Patch-Gen to generate realistic and diverse defect shapes, which narrows the domain gap between training and testing. Our R3D-AD ensures a uniform spatial transformation, which allows straightforwardly generating anomaly results by distance comparison. Extensive experiments show that our R3D-AD outperforms previous state-of-the-art methods, achieving 73.4% Image-level AUROC on the Real3D-AD dataset and 74.9% Image-level AUROC on the Anomaly-ShapeNet dataset with an exceptional efficiency.

Keywords: 3D anomaly detection, industrial applications, 3D reconstruction, self-supervised learning

1 Introduction

Anomaly detection aims to identify instances containing anomalies and to precisely locate the specific positions of defects. This task is extensively applied across multiple fields and plays a crucial role in quality control within industrial production [28]. 3D anomaly detection [19] has emerged due to its intrinsic

* Equal contribution.

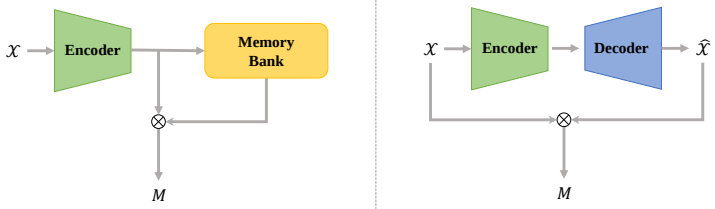

Fig. 1: Comparison of architectures. **(a)** Embedded model encodes the input $\mathcal{X}$ into features and stores them in the memory bank during training. The anomaly map $\mathcal{M}$ is obtained by comparing the test features with all the features in the memory bank. **(b)** Reconstructive model is trained by minimizing the loss between its input $\mathcal{X}$ and the output $\widehat{\mathcal{X}}$. The anomaly map $\mathcal{M}$ is obtained by comparing the test phase input with its corresponding reconstruction target.

modality superior for avoiding blind spots in advanced processing and precision manufacturing. However, the discrete and disordered data form of point clouds makes it more difficult to acquire features compared to images. With the scarcity of anomalies, 3D anomaly detection also faces the problem of domain shift while only normal data are presented during training. The presence of these issues underscores the necessity and urgency of devising an efficient framework for the 3D anomaly detection task.

Similar to traditional 2D anomaly detection [28, 42], current 3D anomaly detection can be primarily categorized into embedding-based and reconstruction-based, as illustrated in Fig. 1. The embedding-based methods involve mapping features extracted with a pre-trained encoder onto a normal distribution for learning. Distributions that do not fall within the interval are classified as anomalies. Most existing 3D anomaly detection methods are based on a memory bank mechanism [3, 11, 19, 36], which stores some representative features during the training phase to implicitly construct a feature distribution. In the testing phase, the presence of anomalies is determined by calculating the Euclidean distance between the input test object and all template point clouds stored in memory. The reconstruction-based methods train a network capable of accurately reconstructing normal point clouds, under the presumption that anomalous point clouds will not be effectively reconstructed since they are not included during training. The anomaly map is produced through the comparison of discrepancies between the input point cloud and its reconstruction. IMRNet [18] employs PointMAE [25] to reconstruct the input in several iterations, getting the final anomaly map by calculating the explicit spatial coordinate differences and implicit deep feature differences of the point cloud, respectively.

However, existing methods face two key issues, high resource cost and irreparable reconstruction. Firstly, methods based on the memory bank [3, 11, 19, 36] store all features from the training phase, each test point cloud needs to be compared with all samples in the memory bank, significantly increasing memory

overhead and inference time costs. This makes such methods almost inapplicable in real industrial production lines due to their inefficiency. Secondly, masked autoencoder (MAE) mechanism [9,25,38] only reconstructs the masked portions of the input, defects within unmasked portions may be preserved. This contradicts the fundamental assumption of detecting anomalies by comparing the original defect-containing point cloud with a reconstructed anomaly-free version. These methods inevitably lead to incorrect reconstructions, undermining their effectiveness in accurately localizing defects.

We propose R3D-AD, a novel 3D anomaly detection method that does not suffer from the space burden and time endurance in memory-based embedded models nor the anomaly unmasking probability in the MAE-based reconstructive models. In contrast to PointMAE, one of our key insights is to perform undifferentiated masking for 3D objects via the noise diffusion mechanism, which maximizes the preservation of anomaly-free shapes and reconstructs abnormal regions. In the reparameterized diffusion process, one-step full mask and reconstruction are achieved by converting the point cloud distribution, instead of the multiple iterative method [18]. We hypothesize that anomaly detection verifies the gap between the reconstructed shapes and the positive samples by learning point movement. Specifically, for input models with arbitrary anomalies, we encode them as latent shape embeddings as decoding conditions and explicitly control the point cloud reconstruction process by step-wise displacements (SWD) decoding. The shape embedding harbors abundant global features and makes it easier to train the network without dwelling on the introduction of local anomaly details. Another key to our approach is to implement a controllable method of point-wise displacement during the diffusion process to refine the point cloud deformation iteratively. We propose to inject latent shape embedding into each step of the inverse denoising process, which drives the anomalous regions to converge to a smooth surface. We further adopted a 3D anomaly simulation strategy Patch-Gen to address the limitations of the dataset, which generates abundant defectives by producing spatial irregularity that is faithful to the real scene, including bulges, sinks, etc. This point cloud data augmentation encourages the self-supervised model to reconstruct more realistic anomaly-free shapes when facing the actual anomaly.

To the best of our knowledge, this is the very first attempt at exploring diffusion in reconstruction-based 3D anomaly detection. Our main contributions are summarized as follows: **(i)** We introduce a novel framework, termed R3D-AD, which performs a one-step full mask and anomaly-free reconstruction for fast and accurate 3D anomaly detection. **(ii)** We propose to learn the step-wise displacement in the reverse diffusion process to explicitly control the reconstruction of anomalous shapes. **(iii)** We introduce a 3D anomaly simulation strategy named Patch-Gen to address the limitation of the data anomaly patterns and improve the reconstruction performance in a supervised setting. **(iv)** Extensive experiments demonstrate that our R3D-AD has achieved state-of-the-art performance on both Real3D-AD and Anomaly-ShapeNet datasets.

2 Related work

2.1 2D Anomaly Detection

Anomaly detection has received increasing attention from researchers in recent years, and many new methods have been proposed to address the problem. Flow-based methods [8, 29, 35, 39] use learned distributions and flow's bijective properties to spot defects, while Memory-based approaches [1, 14, 28] gauge anomaly scores by contrasting test sample features with memory bank-stored norms. Reconstruction-based models [2, 41, 42] flag anomalies by comparing inputs to their online reconstructions. Recent works [12, 16, 32, 43] augment the anomaly detection datasets with generated synthetic anomalies to compensate for the negative example scarcity problem.

2.2 3D Anomaly Detection

This field lags behind the development of 2D anomaly detection since 3D data are harder to obtain, while point cloud data are sparser and contain more noise than image data. BTF [11] integration of handcrafted 3D descriptors with classic 2D method PatchCore [28], constructing a basic framework for 3D anomaly detection. M3DM [36] advances the field by separately analyzing features from point clouds and RGB images, then merging these for improved decision-making. CPMF [3] converts point clouds into two-dimensional images from multiple angles, extracting additional features from these images with a pre-trained network, and enhancing detection capabilities through information fusion. Reg3D-AD [19] develops a registration-based method, the RANSAC algorithm was used to align each sample before comparing it to the stored template during the test phase. IMRNet [18] trains a PointMAE [25] to reconstruct anomaly-free samples and identifies anomalies by juxtaposing the reconstructed point cloud against the initial input. Many of these use memory banks to store the features of the training samples or require multiple iterations to restore points. Unlike previous methods, our approach requires only one step of reconstruction and has significant advantages in both time and space efficiency.

2.3 Diffusion Models

Diffusion models have proven their effectiveness in several generative tasks, such as image generation [31], speech generation [15], and video generation [10]. Denoising Diffusion Probabilistic Models (DDPMs) [13, 33, 34] employ a forward noising mechanism, incrementally integrating Gaussian noise into images, alongside a reverse process meticulously trained to counteract the forward mechanism. Denoise AD [22] conducts DDPM for reconstructing within the features space, generating images that contain less noise. In recent years, many studies [6, 17, 20, 24] have attempted to use the diffusion model to explore the 3D reconstruction task. DPM [23] incorporates a shape latent variable to encapsulate the geometric intricacies of 3D shapes, it distinctively models this variable's

distribution utilizing Normalizing Flows [7, 27]. PVD [44] utilizes PVCNNs [21] for the point-voxel representation of 3D shapes and integrates structured locality into point clouds. This innovative approach leverages the strengths of both point and voxel representations, optimizing the model's ability to capture the intricate spatial hierarchies and local geometries within 3D objects. Since diffusion-based reconstruction recovers the target shape from complete noise, the dilemma of reconstructing only the masked region in the MAE [9] mechanism does not exist.

3 Method

3.1 Overview

We model the anomaly detection problem as mapping an anomalous point cloud $\mathcal{P}_{\mathrm{a}} \in \mathbb{R}^{N\times 3}$ to a positive shape with which it is aligned. The framework of R3D-AD is shown in Fig. 2, where the simulated anomalous shapes are reconstructed in a self-supervised setting in the training phase and then compared with the original input to detect anomalies. The reconstructed anomaly-free model is aligned with the input, thus allowing direct computation of anomaly scores and segmentation of anomalous regions by conditioned distance functions. Simultaneously, the anomaly simulation strategy faithfully generates realistic defects and randomly synthesizes diverse anomaly shapes on normal samples, improving the generalization ability of the network in the case of limited anomaly samples.

3.2 Preliminary of denoising diffusion probabilistic models

A DDPM is inspired by the thermal diffusion process in an evolving thermodynamic system, which consists of a diffusion process and a reverse process.

The forward Markovian process gradually adds Gaussian noise to a clean sample $\boldsymbol{x}^{(0)}$ from a data distribution $q(\boldsymbol{x}^{(0)})$ and turns it into a Gaussian noise $\boldsymbol{x}^{(T)}$, which is defined as

$$q(\boldsymbol{x}^{(0)}, ..., \boldsymbol{x}^{(T)}) = \prod_{t=1}^{T} q(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(t-1)}), \tag{1}$$

where $q(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(t-1)}) = \mathcal{N}(\boldsymbol{x}^{(t)}; \sqrt{1-\beta_t}\boldsymbol{x}^{(t-1)}, \beta_t \boldsymbol{I})$ is the Markov diffusion kernel, $t = 1, ..., T$, T is the number of diffusion steps, and β_t is a variance schedule. We have $q(\boldsymbol{x}^{(t)}|\boldsymbol{x}^{(0)}) = \mathcal{N}(\boldsymbol{x}^{(t)}; \sqrt{\overline{\alpha}_t}\boldsymbol{x}^{(0)}, (1-\overline{\alpha}_t)\boldsymbol{I})$ by reparameterization with $\alpha_t = 1 - \beta_t$, $\overline{\alpha}_t = \prod_{s=0}^{t} \alpha_s$. $\boldsymbol{x}^{(t)}$ can be sampled by

$$\boldsymbol{x}^{(t)} = \sqrt{\overline{\alpha}_t}\boldsymbol{x}^{(0)} + \epsilon\sqrt{(1-\overline{\alpha}_t)}, \tag{2}$$

where $\boldsymbol{\epsilon}$ is a standard Gaussian noise and $\epsilon \sim \mathcal{N}(0, \boldsymbol{I})$. When T is large enough, $\boldsymbol{x}^{(T)}$ will eventually become a Gaussian noise.

The reverse process is also a Markovian process that denoises over a series of steps to generate meaningful data from the target distribution $q(\boldsymbol{x}^{(0)})$. The

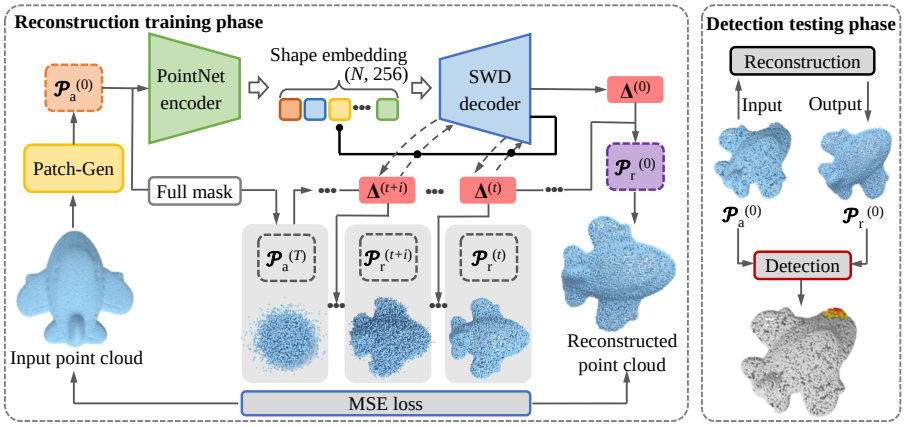

Fig. 2: Overall architecture of R3D-AD for shape reconstruction and anomaly detection of point cloud objects. **Reconstruction training phase**: The simulated anomalous $\mathcal{P}_{\mathrm{a}}^{(0)}$ is generated resort to Patch-Gen from the input point cloud. It further fully masked as $\mathcal{P}_{\mathrm{a}}^{(T)}$ while also encoded to latent shape embedding. The SWD decoder then explicitly reconstructs the anomaly-free object $\mathcal{P}_{\mathrm{r}}^{(0)}$ with consistent spatial transform by conditionally generating point-level displacements $\Delta^{(t)}$ at each step of the inverse process. **Detection testing phase**: The test point cloud $\mathcal{P}_{\mathrm{a}}^{(0)}$ is reconstructed to $\mathcal{P}_{\mathrm{r}}^{(0)}$ with normal shape, and compared at a distance level to detect the anomalous region.

inverse process denoises the noise $\boldsymbol{x}^{(T)}$ from a distribution $p(\boldsymbol{x}^{(T)})$, which is defined as

$$p_\theta(\boldsymbol{x}^{(0)}, ..., \boldsymbol{x}^{(T-1)}|\boldsymbol{x}^{(T)}, \boldsymbol{c}) = \prod_{t=1}^{T} p_\theta(\boldsymbol{x}^{(t-1)}|\boldsymbol{x}^{(t)}, \boldsymbol{c}), \tag{3}$$

where $p_\theta(\boldsymbol{x}^{(t-1)}|\boldsymbol{x}^{(t)}), \boldsymbol{c} = \mathcal{N}(\boldsymbol{x}^{(t-1)}; \mu_\theta(\boldsymbol{x}^{(t)}, t, \boldsymbol{c}), \sigma_t^2 \boldsymbol{I})$, the mean $\boldsymbol{\mu}_\theta(\boldsymbol{x}^{(t)}, t, \boldsymbol{c})$ is estimated by a neural network parameterized by $\boldsymbol{\theta}$, $\boldsymbol{c}$ is the latent condition encoding, and σ_t^2 is a step-dependent variance. $\boldsymbol{\mu}_\theta$ can be reparameterized as

$$\mu_\theta(\boldsymbol{x}^{(t)}, t, \boldsymbol{c}) = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{x}^{(t)} - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}^{(t)}, t, \boldsymbol{c})), \tag{4}$$

where $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}^{(t)}, t, \boldsymbol{c})$ is a neural network utilized to denoise the Gaussian noise from $\boldsymbol{x}^{(T)}$.

The training objective is minimized by training $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}^{(t)}, t, \boldsymbol{c})$ to approximate $\boldsymbol{\epsilon}$. The training objective is defined as

$$\mathcal{L} = \mathbb{E}_{t\sim[1:T], \boldsymbol{x}^{(\mathbf{0})}\sim q(\boldsymbol{x}^{(\mathbf{0})}), \epsilon\sim\mathcal{N}(0,\boldsymbol{I})} \parallel \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\sqrt{\overline{\alpha}_t}\boldsymbol{x}^{(0)} + \sqrt{1-\overline{\alpha}_t}\epsilon, t, \boldsymbol{c}) \parallel, \tag{5}$$

where t is sampled from the uniform distribution over 1,2, ..., T, $q(\boldsymbol{x}^{(\mathbf{0})})$ is the distribution of $\boldsymbol{x}^{(\mathbf{0})}$, and $\boldsymbol{\epsilon}$ is the Gaussian noise.

3.3 Diffusion-based 3D anomaly reconstruction

We formulate the point cloud reconstruction task of the anomaly-free model as the conditional generation, which decodes the explicit displacement with the target distribution $q(\mathcal{P}_{\mathrm{r}}|\mathcal{P}_{\mathrm{a}}^{(T)}, \boldsymbol{c})$, where $\boldsymbol{c}$ is the decoding condition. The essential question of anomaly detection in this paper is how to conditional reconstruct anomaly-free shapes on the reference of input point clouds with different spatial transformations. Since there is a high similarity of global features between abnormal and normal samples during self-supervised reconstruction, the most immediate approach is to extract an efficient global feature from input to serve as an auxiliary conditional embedding for the denoising function $\boldsymbol{\epsilon_\theta}$. We implement the encoding of latent shape embedding $\boldsymbol{c}$ as a conditional input to guide reconstruction in the reverse diffusion process.

Latent shape embedding The feature encoder aims to encode the point cloud to the latent shape embedding $\boldsymbol{c}$ with high-level features for the conditional generation process. Different from other global-local extracting methods [37,40], we focus more on extracting global features, which characterize the semantic information of shape and pose of most anomaly-free regions in the point cloud. The feature encoder mainly consists of cascaded multi-layer perceptions (MLP) based on PointNet [5]. It implements max-pooling after mapping $\mathcal{P}_{\mathrm{a}}^{(0)}$ to different dimensions and then compresses them to extract the global shape embedding.

Step-wise displacement decoding To achieve point cloud reconstruction with transformation consistency while preserving the structure of non-anomalous regions, our method injects latent shape embedding $\boldsymbol{c}$ to the decoder at each step of the reverse diffusion process, as shown in Fig. 2. In principle, in the training phase, $\boldsymbol{\epsilon_\theta}$ learns the added Gaussian noise in the forward diffusion process by the decoder to model the conditional probability distribution. Conditionally generating target shapes from $N \times 3$ Gaussian noise is a straightforward approach, but it is afflicted by the issues of reconstructing the point cloud details and transform consistency. Learning the relative deformation of points for anomalous objects is more efficient. Considering the mapping degradation of the vanilla autoencoder in the reconstruction training phase [22], we utilize the Gaussian noise of the forward process Eq. 2 to fully mask the point cloud object directly without blind spots, preventing the decoding process from receiving negative state shapes. The masked points $\mathcal{P}_{\mathrm{a}}^{(T)}$ and latent shape embedding $\boldsymbol{c} \in \mathbb{R}^{256}$ are as the inputs of the SWD decoder. The point-wise displacement vector $\Delta^{(t)}$ is generated at each step of the iterative process thus disentangling the prediction noise and the desired anomaly-free shape. The reverse process can be defined according to Eq. 3 and the displacement vector $\Delta^{(t)}$ can be represented by

$$\Delta^{(t-1)} = \frac{1}{\sqrt{\alpha_t}}(\Delta^{(t)} - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\boldsymbol{\epsilon_\theta}(\Delta^{(t)}, \beta_t, \boldsymbol{c})) + \sigma\boldsymbol{\epsilon}, \tag{6}$$

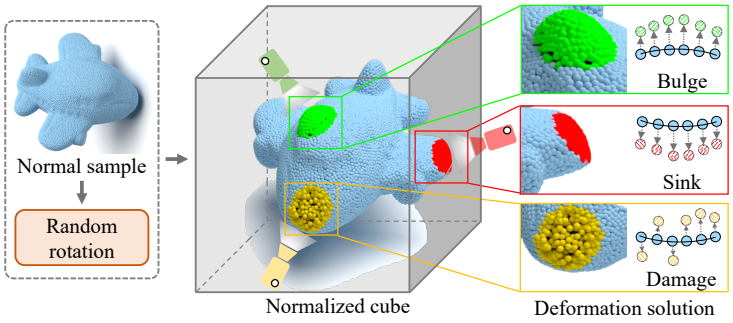

Fig. 3: Illustration of Patch-Gen, the 3D anomaly simulation strategy. The input normal point cloud is first randomly rotated. On the surface of the normalized cube, we randomly select viewpoints to find the nearest patch of points. The selected points are then transformed into irregular defects according to the specific deformation solution.

where σ is the variance. A PointwiseNet is adopted for $\boldsymbol{\epsilon_\theta}$ to decode the $\Delta^{(t-1)}$ from the previous step and $\boldsymbol{c}$. β_t is used to generate trigonometric position embedding $\boldsymbol{e}_\mathrm{p} = (\beta_t, \sin(\beta_t), \cos(\beta_t))$. $\boldsymbol{e}_\mathrm{p}$ is concatenated with $\boldsymbol{c}$ and then fed into the concatenate-squash linear module of PointwiseNet with a residual function. The output reconstructed point cloud at the t step is $\mathcal{P}_\mathrm{r}^{(t)} = \mathcal{P}_\mathrm{r}^{(t+1)} + \Delta^{(t)}$. The registered original and reconstructed objects are distinguished from the anomalous shape by the anomaly scores based on the conditioned distance function.

3.4 3D anomaly simulation strategy

Given that a small number of normal samples is not conducive for the model to learn diverse and essential features, we propose the Patch-Gen strategy to simulate the defects from anomaly-free shapes for training data augmentation. Patch-Gen encourages the reconstruction model to learn to detect irregularity, where the anomaly-free point clouds and their diverse anomaly patterns are integrated into training pairs and are utilized to learn the discrimination feature between normal and anomalous surfaces. The intuition is that the diversity of simulated negative samples forces our network to learn how to reconstruct anomaly-free shapes instead of memorizing their complete outfits.

As shown in Fig. 3, the input normal sample is first randomly rotated. The random spatial rotation is designed to improve the generalization capability for test samples with very different spatial transformations, as defined by:

$$\mathcal{P}_\mathrm{a} = \mathcal{P} \cdot \mathcal{R}, \tag{7}$$

where $\mathcal{P}$ is input normal sample point cloud and $\mathcal{R} \in \mathbb{R}^{3\times3}$ is obtained by randomly selecting rotation angles for all three axes. In addition to global shape awareness of the model by the random rotation, we further perform a fine granularity of the anomaly simulation. We randomly take a viewpoint $\mathcal{P}_v$ from the surface of the cube. Therefore, the patch of nearest N points $\mathcal{P}_n$ from $\mathcal{P}_a$ can be

determined according to the $\mathcal{P}_v$. The shape augmentation scheme Patch-Gen is defined as follows:

$$\mathcal{P}_n = \mathcal{P}_n + S \cdot \mathit{normalize}(\mathcal{P}_n - \mathcal{P}_v) \odot \mathcal{T}, \tag{8}$$

where *nomalize* represents a normalization operation on a vector, S is a predefined hyper-parameter that controls the scaling of the patch points, and $\mathcal{T}$ is the translation matrix originating from a Gaussian distribution. The $\mathcal{P}_{\mathrm{a}}$ is finally obtained by only updating the patch region $\mathcal{P}_n$.

With the proposed Patch-Gen, we can simulate the generation of multiple anomalies, which is mainly done by controlling $\mathcal{T}$. Bulge or sink can be generated by sorting $\mathcal{T}$ after sampling from the distribution, while damage can be generated by direct overlaying without manipulation.

3.5 Training objective

In the reconstruction task of the object with N points, the network learns a diffusion model with an $\mathbb{R}^{N\times 3} \rightarrow \mathbb{R}^{N\times 3}$ mapping relation. Iterative denoising under the semantic condition of point embedding realizes the prediction of point offsets. Concretely, the network is trained to learn the noise that needs to be eliminated to recover the anomaly-free shape with the L_2 distance between the ground truth and the denoised reconstructed points. We make use of the mean squared error (MSE) loss as the primary reconstruction loss which evaluates the mean squared error of the element-wise distances between $\mathcal{P}_{\mathrm{a}}^{(0)}$ and $\mathcal{P}_{\mathrm{r}}^{(0)}$. The MSE training loss is formulated as:

$$\mathcal{L}_{\mathcal{P}_{\mathrm{a}},\mathcal{P}_{\mathrm{r}}} = \frac{1}{N}\sum_{i=1}^{N} {}_{p_{\mathrm{a}}\in\mathcal{P}_{\mathrm{a}},p_{\mathrm{r}}\in\mathcal{P}_{\mathrm{r}}} \parallel p_{\mathrm{a}} - p_{\mathrm{r}} \parallel^2. \tag{9}$$

4 Experiments

4.1 Datasets

Real3D-AD [19] is a 3D anomaly detection dataset based on real samples, exhibiting a higher point precision and spatial distance per point cloud. Each category contains 4 training samples and 100 test samples. The training set contains 360° complete surface point clouds of the objects, which are obtained by manually calibrating and stitching the scans of multiple sides of the objects. The test samples are scans only one side with a huge difference from the training set. The distribution of the point clouds also varies among the total 12 categories, further deepening the detection difficulty compared to 2D scenes.

Anomaly-ShapeNet [18] is a 3D anomaly detection, crafted through modifications to the synthetic samples found in ShapeNetCorev2 [4]. It contains 40 diverse categories, featuring over 1600 samples of its complete surface point

Method	BTF [11]		M3DM [36]	PatchCore [28]		CPMF [3]	Reg3D-AD [19]	IMRNet [18]	**Ours**
Feat.	*Raw*	*FPFH*	*PointMAE*	*FPFH*	*PointMAE*	*ResNet*	*PointMAE*	*PointMAE*	*Raw*
Airplane	0.730	0.520	0.434	**0.882**	0.726	0.701	0.716	0.762	<u>0.772</u>
Candybar	0.539	0.630	0.552	0.541	0.663	0.552	0.685	**0.755**	<u>0.696</u>
Car	0.647	0.560	0.541	0.590	0.498	0.551	0.697	<u>0.711</u>	**0.713**
Chicken	0.789	0.432	0.683	<u>0.837</u>	0.827	0.504	**0.852**	0.780	0.714
Diamond	0.707	0.545	0.602	0.574	0.783	0.523	<u>0.900</u>	**0.905**	0.685
Duck	0.691	<u>0.784</u>	0.433	0.546	0.489	0.582	0.584	0.517	**0.909**
Fish	0.602	0.549	0.540	0.675	0.630	0.558	**0.915**	<u>0.880</u>	0.692
Gemstone	**0.686**	0.648	0.644	0.370	0.374	0.589	0.417	<u>0.674</u>	0.665
Seahorse	0.596	**0.779**	0.495	0.505	0.539	<u>0.729</u>	0.762	0.604	0.720
Shell	0.396	<u>0.754</u>	0.694	0.589	0.501	0.653	0.583	0.665	**0.840**
Starfish	0.530	0.575	0.551	0.441	0.519	<u>0.700</u>	0.506	0.674	**0.701**
Toffees	0.703	0.462	0.450	0.565	0.585	0.390	**0.827**	<u>0.774</u>	0.703
Average	0.635	0.603	0.552	0.593	0.595	0.586	0.704	<u>0.725</u>	**0.734**

Table 1: Image-level anomaly detection AUROC on Real3D-AD dataset. We highlight the best result in **bold** and the second best result with an <u>underline</u>.

clouds. Each category's training set contains merely 4 samples, while the test sets are designed to assess the model's performance across both normal and a spectrum of abnormal samples. It widely increases the anomaly types while keeping the number of points the same as the previous studies, which places higher demands on the robustness and generality of the proposed algorithms.

4.2 Evaluation metrics

For image-level anomaly detection, the Area Under the Receiver Operating Curve (AUROC) is utilized in line with established practices. For the evaluation of pixel-level anomalies, the AUROC metric is similarly applied in the context of point segmentation accuracy. A value of 0.5 of the AUROC score denotes no discriminative capability (equivalent to random guessing), whereas a score of 1.0 indicates perfect discrimination between positive and negative classes.

4.3 Implementation details

Our methodology is implemented using PyTorch [26] with end-to-end training across the network. The optimization is performed using the Adam optimizer, starting at an initial learning of 0.001. The training process involves a total batch size of 128 across 40,000 iterations for comprehensive learning. All input point clouds undergo a preprocessing step where they are randomly downsampled to a fixed size of 4096 and 2048 points on Real3D-AD and Anomaly-ShapeNet, respectively. Additionally, we normalized these point clouds by setting their center of gravity as the origin of coordinates and scaling their dimensions to fall within the range of -1 to 1, optimizing for the diffusion process.

Method	BTF [11]		M3DM [36]	PatchCore [28]		CPMF [3]	Reg3D-AD [19]	IMRNet [18]	**Ours**
Feat.	*Raw*	*FPFH*	*PointMAE*	*FPFH*	*PointMAE*	*ResNet*	*PointMAE*	*PointMAE*	*Raw*
Ashtray	0.578	0.420	0.577	0.587	0.591	0.353	0.597	<u>0.671</u>	**0.833**
Bag	0.410	0.546	0.537	0.571	0.601	0.643	<u>0.706</u>	0.660	**0.719**
Bottle	0.558	0.404	0.584	0.614	0.588	0.469	0.569	<u>0.631</u>	**0.750**
Bowl	0.470	0.581	0.579	0.558	0.547	<u>0.679</u>	0.548	0.676	**0.751**
Bucket	0.469	0.517	0.405	0.510	0.577	0.542	<u>0.681</u>	0.676	**0.719**
Cap	0.509	0.562	0.599	0.645	0.583	0.601	0.632	<u>0.704</u>	**0.726**
Cup	0.462	0.598	0.548	0.593	0.583	0.498	0.524	<u>0.700</u>	**0.767**
Eraser	0.525	<u>0.719</u>	0.627	0.657	0.677	0.689	0.343	0.548	**0.890**
Headset	0.447	0.505	0.597	0.610	0.609	0.551	0.574	<u>0.698</u>	**0.767**
Helmet	0.508	0.569	0.488	0.465	0.495	0.532	0.491	<u>0.603</u>	**0.704**
Jar	0.420	0.424	0.441	0.472	0.483	0.610	0.592	<u>0.780</u>	**0.838**
Microphone	0.563	0.671	0.357	0.388	0.488	0.509	0.414	<u>0.755</u>	**0.762**
Shelf	0.164	0.609	0.564	0.494	0.523	0.685	<u>0.688</u>	0.603	**0.696**
Tap	0.549	0.553	0.747	<u>0.760</u>	0.498	0.528	0.659	0.686	**0.818**
Vase	0.517	0.464	0.534	0.554	0.582	0.514	0.576	<u>0.629</u>	**0.734**
Average	0.493	0.528	0.552	0.568	0.562	0.559	0.572	<u>0.659</u>	**0.749**

Table 2: Image-level anomaly detection AUROC on Anomaly-ShapeNet dataset. We highlight the best result in **bold** and the second best result with an <u>underline</u>.

4.4 Main results

We conduct experiments on Real3D-AD [19] based on real sampling and Anomaly-ShapeNet [18] based on simulation.

As shown in Table 1, we first compare the image-level AUROC metric with current cutting-edge 3D anomaly detection models on Real3D-AD. It shows that our method achieves the best performance using only raw point cloud data, while most of the existing methods use Fast Point Feature Histograms (FPFH) operator [30] or ShapeNet [4] pre-trained PointMAE [25] as feature extractor. Due to significant disparities in quantity, size, and distribution among different categories of point clouds in Real3D-AD, scoring variations across categories are more pronounced with other methods. For instance, numerous methods perform under 0.5 in certain categories, indicating their inadequacy in extracting meaningful features while facing challenging samples. In contrast, our method not only exhibits superior performance in 3D anomaly detection across the majority of categories but also achieves the best overall average across all categories. This demonstrates the strong generalizability and robustness of our approach.

We further evaluate our method on Anomaly-ShapeNet in Table 2, which encompasses a broader array of categories and a greater diversity of defect types. Compared to Real3D-AD, Anomaly-ShapeNet significantly enhances the diversity of defects, wherein the increased variety of defect types further escalates the complexity of detection tasks. The results highlight the exceptional performance of our method across all evaluated categories, demonstrating an average improvement of 9% on AUROC relative to the approaches previously utilized.

Model	Diffusion	Condition	Relative	Patch-Gen	I-AUROC	P-AUROC
A	✓	✗	✗	✗	0.586	0.524
B	✓	✓	✗	✗	0.667	0.513
C	✓	✓	✓	✗	0.712	0.573
D	✓	✓	✓	✓	**0.734**	**0.592**

Table 3: Ablation studies for 3D adaptation components on Real3D-AD dataset.

4.5 Ablation study

To delve into the effect of individual components, we conduct ablation experiments on the Real3D-AD dataset. To fully demonstrate and compare the performance of the models, we report both image-level and pixel-level results with I-AUROC and P-AUROC, respectively.

Main component Table 3 compares the performance of different variants from R3D-AD, which includes the influence of the denoising condition embedding, displacement-based reconstruction way, and the data augmentation strategy of Patch-Gen. Model A is denoted as our baseline, which is a vanilla DDPM model for point cloud reconstruction. Introducing a condition into the DDPM (Model B) significantly boosts performance, particularly in terms of I-AUROC, which sees a 13.8% increase to 0.667. Model C, which predicts point displacements based on conditional DDPM, preserving detailed structural information while accommodating the relative displacement of points contributes to a notable 6.0% gain in P-AUROC over Model B. Model D is trained under the conditions of shape embedding with the Patch-Gen strategy. Considering that the defective portion contains only a small portion of the original point cloud, we try to reconstruct the relative displacement in a way that preserves as much detail as possible, which is effective for both 3D anomaly detection and segmentation.

Patch-Gen Table. 4 analyzes the influence of two key parameters in Patch-Gen: the selection points ratio and the scaling points factor.

The **selection points ratio** from Table. 4a determines the proportion of points in the point cloud that are selected for transformation. Our findings suggest that a selection ratio of 1/32 achieves the best performance. It appears that this ratio provides a balanced trade-off between maintaining sufficient structure for anomaly detection and introducing enough variation to simulate anomalies effectively. Notably, as the ratio increases beyond 1/16, both I-AUROC and P-AUROC scores decrease in severity, since real defects only account for a small portion of the overall point cloud, a wide selection of points not only destroys the structure of the original point cloud, but also makes the distribution of the training and test sets inconsistent.

The **scaling points factor** is the intensity of the random transformation applied to the selected points, as detailed in Table 4b. The optimal performance is

ratio	I-AUROC	P-AUROC
1/64	0.716	0.584
1/32	**0.734**	**0.592**
1/16	0.727	0.579
1/8	0.683	0.528

(a) Selection points ratio.

factor	I-AUROC	P-AUROC
0.1	**0.734**	**0.592**
0.2	0.727	0.572
0.4	0.715	0.554
0.8	0.661	0.517

(b) Scaling points factor.

Table 4: Ablation studies for Patch-Gen implementation on Real3D-AD dataset. Default settings are marked in gray.

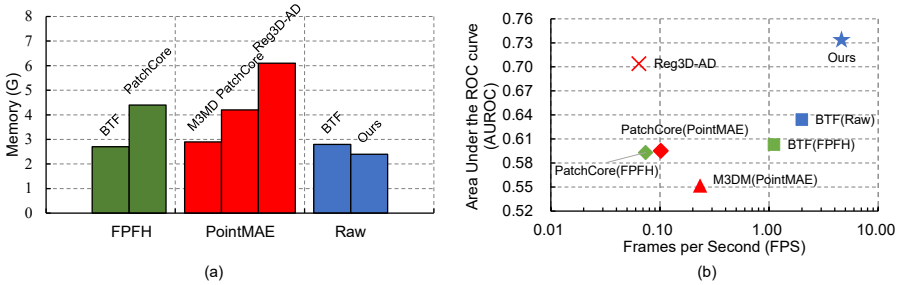

Fig. 4: Memory and time cost during inference on Real3D-AD dataset. **(a)** Memory usage comparison between different models. **(b)** 3D anomaly detection performance vs. frames per second on an NVIDIA RTX 3090 GPU. Our R3D-AD outperforms all previous methods on both accuracy and efficiency by a significant margin.

observed at a scaling factor of 0.1, which implies that minor transformations are more effective for simulating anomalies without significantly altering the original data distribution. Larger scaling factors lead to a consistent decline in performance, underscoring the importance of subtle transformations for preserving the utility of the simulated anomalies for detection tasks.

Memory and time cost As depicted in Figure 4, we evaluate the disparity in both storage consumption and inference time of our model under identical experimental conditions, compared to existing methods. Regarding memory usage, our approach demonstrates a marked superiority by employing raw coordinate features instead of FPFH or PointMAE features, significantly reducing the memory footprint. Since no memory bank exists, our method is also more space-efficient compared to BTF which also uses raw features. Moreover, our method eliminates the necessity to compare all the features in memory, substantially increasing operational efficiency. The implementation of Patch-Gen inherently bestows our model with exceptional robustness, enabling precise reconstruction of point clouds from various angles without the need for the time-intensive RANSAC alignment process required by Reg3D-AD.

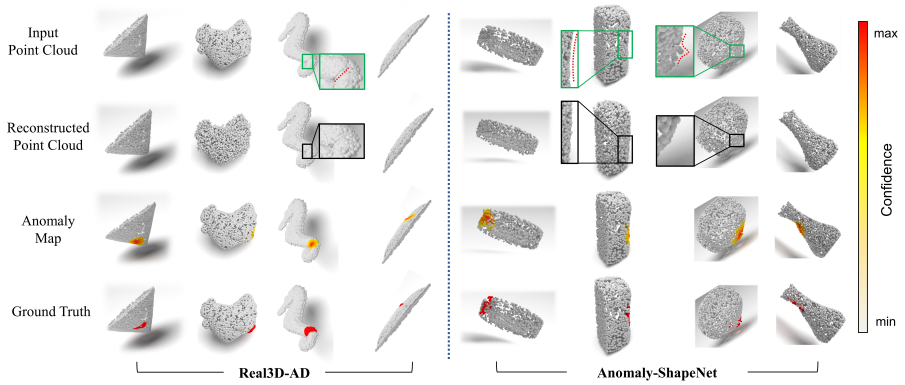

Fig. 5: Qualitative analysis on Real3D-AD dataset and Anomaly-ShapeNet dataset. The anomaly map is obtained directly by calculating the differences between the input and reconstructed point clouds, where deeper colors represent more confidence.

4.6 Qualitative results

Figure 5 presents some qualitative outcomes, with varying shades of color indicating different levels of anomaly scores. We select several representative defective samples to demonstrate the robustness of our algorithm. The left four columns display samples from Real3D-AD, while the right four columns samples are from Anomaly-ShapeNet. The illustration reveals that our R3D-AD algorithm has precisely reconstructed the defective portions of the point cloud across various samples: the deep *sink* in the Seahorse sample, the *concavity* in the Bag sample, and the *bulge* in the Jar sample. Leveraging the accurately reconstructed point clouds, final point cloud segmentation maps are also produced, further evidencing the efficacy of our approach.

5 Conclusion

In this work, we presented R3D-AD, a novel reconstructive 3D anomaly detection model based on conditional diffusion. Our goal is to overcome the limitations faced by current 3D anomaly detection methods, such as the inefficiencies due to the memory bank module and low performance caused by incorrect rebuilds with MAE. To address these challenges, we leverage the diffusion process for full reconstruction, followed by a direct comparison between the input and the reconstructed point cloud to obtain the final anomaly score. The embedded latent variable that spans the decoding process, step-wisely generating point-level displacements from the noise to the target anomaly-free sample. We also propose Patch-Gen, a data augmentation tailored for point cloud anomaly simulation. Extensive experiments conducted on 3D anomaly benchmarks validate the superiority of our R3D-AD in comparison to state-of-the-art alternatives in terms of both accuracy and versatility.

Acknowledgements

This work was supported in part by the Pioneer and Leading Goose R&D Program of Zhejiang (Grant No. 2022C01051), in part by the National Natural Science Foundation of China (Grant No. 52375271, 52275274), and in part by the Natural Science Foundation of Zhejiang Province (Grant No. LY23E050011).

References

1. Bae, J., Lee, J.H., Kim, S.: Pni: Industrial anomaly detection using position and neighborhood information. In: ICCV (2023)
2. Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In: VISIGRAPP (2019)
3. Cao, Y., Xu, X., Shen, W.: Complementary pseudo multimodal feature for point cloud anomaly detection. arXiv preprint (2023)
4. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. arXiv preprint (2015)
5. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: CVPR (2017)
6. Chu, R., Xie, E., Mo, S., Li, Z., Nießner, M., Fu, C.W., Jia, J.: Diffcomplete: Diffusion-based generative 3d shape completion. In: NeurIPS (2023)
7. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. In: ICLR (2017)
8. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: WACV (2022)
9. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: CVPR (2022)
10. Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D.P., Poole, B., Norouzi, M., Fleet, D.J., Salimans, T.: Imagen video: High definition video generation with diffusion models (2022)
11. Horwitz, E., Hoshen, Y.: Back to the feature: Classical 3d features are (almost) all you need for 3d anomaly detection. In: CVPRW (2023)
12. Hu, T., Zhang, J., Yi, R., Du, Y., Chen, X., Liu, L., Wang, Y., Wang, C.: Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In: AAAI (2024)
13. Jonathan Ho, Ajay Jain, and Pieter Abbeel: Denoising diffusion probabilistic models. In: NeurIPS (2020)
14. Kim, D., Park, C., Cho, S., Lee, S.: Fapm: Fast adaptive patch memory for real-time industrial anomaly detection. In: ICASSP (2023)
15. Kong, Z., Ping, W., Huang, J., Zhao, K., , Catanzaro, B.: Diffwave: A versatile diffusion model for audio synthesis. In: ICLR (2021)
16. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In: CVPR (2021)
17. Li, M., Duan, Y., Zhou, J., Lu, J.: Diffusion-sdf: Text-to-shape via voxelized diffusion. In: CVPR (2023)
18. Li, W., Xu, X., Gu, Y., Zheng, B., Gao, S., Wu, Y.: Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. arXiv preprint (2023)

19. Liu, J., Xie, G., Li, X., Wang, J., Liu, Y., Wang, C., Zheng, F., et al.: Real3d-ad: A dataset of point cloud anomaly detection. In: NeurIPS (2023)
20. Liu, Z., Feng, Y., Black, M.J., Nowrouzezahrai, D., Paull, L., Liu, W.: Meshdiffusion: Score-based generative 3d mesh modeling. In: ICLR (2023)
21. Liu, Z., Tang, H., Lin, Y., Han, S.: Point-voxel cnn for efficient 3d deep learning. In: NeurIPS (2019)
22. Lu, F., Yao, X., Fu, C., Jia, J.: Removing anomalies as noises for industrial defect localization. In: ICCV (2023)
23. Luo, S., Hu, W.: Diffusion probabilistic models for 3d point cloud generation. In: CVPR (2021)
24. Mo, S., Xie, E., Chu, R., Hong, L., Nießner, M., Li, Z.: Dit-3d: Exploring plain diffusion transformers for 3d shape generation. In: NeurIPS (2023)
25. Pang, Y., Wang, W., Tay, F.E., Liu, W., Tian, Y., Yuan, L.: Masked autoencoders for point cloud self-supervised learning. In: ECCV (2022)
26. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. NeurIPS (2019)
27. Rezende, D.J., Mohamed, S.: Variational inference with normalizing flows. In: ICML (2015)
28. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: CVPR (2022)
29. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: Semi-supervised defect detection with normalizing flows. In: WACV (2021)
30. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: ICRA (2009)
31. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S.K.S., Ayan, B.K., Mahdavi, S.S., Lopes, R.G., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M.: Photorealistic text-to-image diffusion models with deep language understanding. arXiv preprint (2022)
32. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for self-supervised anomaly detection and localization. In: ECCV (2022)
33. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint (2020)
34. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: ICLR (2021)
35. Tailanian, M., Pardo, Á., Musé, P.: U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold. arXiv preprint (2022)
36. Wang, Y., Peng, J., Zhang, J., Yi, R., Wang, Y., Wang, C.: Multimodal industrial anomaly detection via hybrid fusion. In: CVPR (2023)
37. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. TOG (2019)
38. Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.: Simmim: A simple framework for masked image modeling. In: CVPR (2022)
39. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. arXiv preprint (2021)
40. Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J., Lu, J.: Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In: CVPR (2022)
41. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: ICCV (2021)

42. Zavrtanik, V., Kristan, M., Skočaj, D.: Reconstruction by inpainting for visual anomaly detection. Pattern Recognition (2021)
43. Zhang, X., Xu, M., Zhou, X.: Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In: CVPR (2024)
44. Zhou, L., Du, Y., Wu, J.: 3d shape generation and completion through point-voxel diffusion. In: ICCV (2021)