

C3D-AD: Toward Continual 3D Anomaly Detection via Kernel Attention with Learnable Advisor

Haoquan Lu¹, Hanzhe Liang^{2,3}, Jie Zhang⁴, Chenxi Hu², Jinbao Wang^{5,6}, Can Gao^{1,6}

¹College of Computer Science and Software Engineering, Shenzhen University

²Shenzhen Audencia Financial Technology Institute, Shenzhen University

³Ningbo Institute of Digital Twin, Eastern Institute of Technology

⁴Faculty of Applied Sciences, Macao Polytechnic University

⁵School of Artificial Intelligence, Shenzhen University

⁶Guangdong Provincial Key Laboratory of Intelligent Information Processing

Abstract

3D Anomaly Detection (AD) has shown great potential in detecting anomalies or defects of high-precision industrial products. However, existing methods are typically trained in a class-specific manner and also lack the capability of learning from emerging classes. In this study, we proposed a continual learning framework named **Continual 3D Anomaly Detection (C3D-AD)**, which can not only learn generalized representations for multi-class point clouds but also handle new classes emerging over time. Specifically, in the feature extraction module, to extract generalized local features from diverse product types of different tasks efficiently, Kernel Attention with random feature Layer (KAL) is introduced, which normalizes the feature space. Then, to reconstruct data correctly and continually, an efficient Kernel Attention with learnable Advisor (KAA) mechanism is proposed, which learns the information from new categories while discarding redundant old information within both the encoder and decoder. Finally, to keep the representation consistency over tasks, a Reconstruction with Parameter Perturbation (RPP) module is proposed by designing a representation rehearsal loss function, which ensures that the model remembers previous category information and returns category-adaptive representation. The proposed method is the first attempt to address 3D anomaly detection in a class-incremental manner, providing the capabilities of multi-class and continual anomaly detection. Extensive experiments on three public datasets demonstrate the effectiveness of the proposed method, achieving an average performance of 66.4%, 83.1%, and 63.4% AU-ROC on Real3D-AD, Anomaly-ShapeNet, and MulSen-AD, respectively.

Introduction

3D Anomaly Detection has garnered significant attention in identifying industrial product defects (Tu et al. 2025; Rudolph et al. 2023). To detect anomalies effectively, feature-embedding methods and reconstruction-based methods have been proposed in recent years (Liu et al. 2024b). These methods extract features that are highly relevant for anomaly detection (Horwitz and Hoshen 2023; Tu et al. 2025; Liang et al. 2025a), leverage complementary modality information to enhance performance (Rudolph et al. 2023;

We have provided the code for C3D-AD with checkpoints and BASELINE at this link: <https://github.com/hzzzzzhappy/CL3AD>.

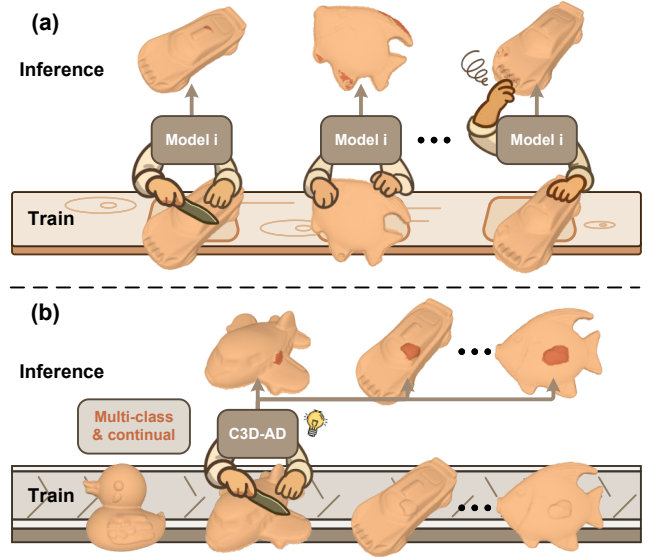


Figure 1: Difference between class-specific model and C3D-AD. (a) Single-class anomaly detection. (b) Multi-class and continual anomaly detection.

Gu et al. 2024), reconstruct point cloud to allow greater generalization across data distribution (Chen et al. 2023; Zhou et al. 2025), and enhance the capability of multi-class anomaly detection (Wang et al. 2025; Cheng et al. 2025). The performance of anomaly detection methods has steadily improved, and the application scenarios have become increasingly close to reality.

However, real-world environments present a more complex challenge, e.g., the continuous emergence of new object categories requires detection. In this context, class-specific models are inefficient as they necessitate complete retraining for each new class. Unified models, while more efficient, are susceptible to catastrophic forgetting, causing the degradation of performance on previously learned tasks. Hence, strategies to mitigate this problem are crucial, which leads to the paradigm of Continual Learning (CL). CL aims to develop models able to learn sequentially from new data without forgetting previously acquired knowledge, as shown in

Figure 1.

Continual learning has been applied to 2D anomaly detection. The rehearsal-based methods, e.g., CAD (Li et al. 2022) and ReplayCAD (Hu et al. 2025), store the information of the Gaussian distribution and parameter of the diffusion model, respectively. The regularization-based methods constrain the parameters to be less sensitive to the new training data. For example, CDAD (Li et al. 2025b) constrains the gradients orthogonal to previous feature representations. Tang et al. (Tang et al. 2025) used a semantic compression strategy to maximize the space margin from different tasks. However, due to high-resolution inputs and class-specific models, the methods based on continual learning cannot be directly applied to 3D point clouds.

Motivated by the observations above, we propose Continual 3D Anomaly Detection (C3DAD), a novel framework designed to address the challenges of sequential anomaly detection in point cloud data. Specifically, we introduce a Kernel Attention with random feature Layer (KAL) to extract generalized features. Rather than directly encoding raw point-level information, KAL extracts the spatial context of the point cloud and mines local structure in the unified kernel space. To efficiently preserve and update multi-class data cache across tasks and reconstruct data correctly and continually, we further propose a Kernel Attention with learnable Advisor (KAA) mechanism for the Encoder-Decoder module. Hence, the module can learn the information from new categories and discard redundant information. Moreover, to mitigate catastrophic forgetting and keep the representation consistent, we proposed the Reconstruction with Parameter Perturbation (RPP) module. This module encourages the model returns category-adaptive representations across all sequential tasks. The main contributions are as follows:

- We introduce a novel layer named KAL to normalize the feature space while extracting features. Leveraging KAL, features are extracted from the unified kernel space, which is significant to continual 3D AD. By extracting local structure information, our method significantly enhances the model’s ability for continual learning.
- We proposed a novel network KAA for continual learning. To address the limitations of fixed-capacity networks in continual learning, we propose a novel encoder-decoder architecture with learnable advisors that reduces redundant information from previous knowledge while learning new knowledge. Traditional attention mechanisms often encounter $\mathcal{O}(n^2)$ complexity, leading to computational bottlenecks. To overcome it, KAA with linear $\mathcal{O}(n)$ complexity updates advisors without compromising effectiveness.
- We present a new hypothesis constraint in continual learning. The network in its current state should retain satisfactory performance on past data when revisited. To enforce this, the network’s future outputs are predicted and constrained between current outputs by RPP, preserving the model’s continual learning abilities over time.

Related Work

3D Anomaly Detection

3D anomaly detection is a computer vision task focused on detecting and scoring anomalous points within 3D data, such as point clouds (Liu et al. 2023; Li et al. 2024, 2025a), to identify product defects in industrial manufacturing. This is achieved through two main methods.

(1) **Feature-embedding methods** (Horwitz and Hoshen 2023; Kruse et al. 2024; Liang et al. 2025a) extract embeddings from 3D data and measure similarity to normal data. Specifically, Student-Teacher networks (Rudolph et al. 2023; Bergmann and Sattlegger 2023; Qin et al. 2023; Gu et al. 2024) effectively assess output differences to indicate anomaly levels. In contrast, (2) **reconstruction-based methods** evaluate errors by comparing outputs to original inputs (Chen et al. 2023; Liang et al. 2025b), allowing greater generalization across data distributions. Researchers have explored various reconstruction frameworks; for example, Masuda et al. (Masuda et al. 2021) proposed an unsupervised anomaly detection framework based on VAE (Kingma and Welling 2022), and Chen et al. (Chen et al. 2023) introduced a novel encoder-decoder for multi-scale and multimodal data. R3D-AD (Zhou et al. 2025), based on diffusion (Ho, Jain, and Abbeel 2020), obscures anomalous geometry for global anomaly detection. Recently, the unified model MC3D-AD (Cheng et al. 2025) demonstrated significant performance improvements for multi-class data, highlighting the value of a one-for-all approach.

However, existing methods struggle to generalize in class-increasing settings. To address this, we propose a novel continual learning framework (C3D-AD) that enables multi-class and continual anomaly detection.

Continual Learning

Continual learning (CL) is a machine learning paradigm that enables models to learn continuously from evolving data streams, adapting to dynamic scenarios without full retraining. It provides lifelong learning capabilities while addressing catastrophic forgetting through three main methods (Mallya, Davis, and Lazebnik 2018). (1) **Regularization-based methods** constrain parameter updates to preserve crucial knowledge (Rebuffi et al. 2017). For instance, Elastic Weight Consolidation uses the Fisher Information Matrix to identify critical parameters, while Learning without Forgetting employs knowledge distillation (Batra and Clark 2024). The STAR method constrains gradient updates using buffered samples (Eskandar et al. 2025). (2) **Rehearsal-based methods** store and reuse past samples, effectively mitigating catastrophic forgetting (Riemer et al. 2019). LiDER enhances network smoothness by optimizing Lipschitz constants (Bonicelli et al. 2022). (3) **Architecture-based methods** allocate specific parameters for each task (Mallya and Lazebnik 2018). The LPS algorithm by Wang et al. partitions the network into task-specific sections to retain information from new tasks (Wang et al. 2020).

CL has been successfully applied to 2D anomaly detection (Barusco et al. 2025). For example, Li et al. (Li

et al. 2025b) proposed CDAD, which projects gradients into a subspace orthogonal to previous feature representations, while Tang et al. (Tang et al. 2025) used a semantic compression strategy to retain essential memories. CAD (Li et al. 2022) and ReplayCAD (Hu et al. 2025) utilize rehearsals of statistical information from previous distributions. Additionally, Liu et al. introduced a continual prompt module in UCAD (Liu et al. 2024a) for task adaptation.

However, due to high-resolution inputs and class-specific models, these methods cannot be directly applied to 3D data, necessitating retraining when encountering new categories. The development of a unified model for 3D anomaly detection in a class-incremental manner has yet to be explored. Therefore, we propose C3D-AD to facilitate the reconstruction of multi-class data and the continual detection of anomalies.

Method

Problem Statement

The task of continual 3D AD is to find a model to detect data from new categories while avoiding catastrophic forgetting. Considering the real application, the model can access only the training data of the current task and memory bank, precluding the storage or revisiting of full data from previous tasks. Hence, the available data arrives sequentially in multiple tasks, i.e., $\mathcal{P}_{\text{train}} = \mathcal{P}_{\text{train}}^1 \cup \mathcal{P}_{\text{train}}^2 \cup \dots \cup \mathcal{P}_{\text{train}}^T$ and $\mathcal{P}_{\text{train}}^i \cap \mathcal{P}_{\text{train}}^j = \emptyset$ ($i \neq j$), where $\mathcal{P}_{\text{train}}^t$ denotes the training data from the t -th task containing point cloud samples from specific categories and only having normal samples, and T is the total number of tasks. In the testing phase, the data to be detected includes both normal and anomalous point cloud samples from all encountered tasks, i.e., $\mathcal{P}_{\text{test}}^1 \subset \mathcal{P}_{\text{test}}^2 \subset \dots \subset \mathcal{P}_{\text{test}}^T = \mathcal{P}_{\text{test}}$. Our objective is to train a unified model that can detect anomalies across all encountered tasks while avoiding catastrophic forgetting of previously learned detection capabilities, which minimizes the cost of training the model.

Preliminary

In this paper, vectors, matrices, and sets are denoted by the bold lowercase letters (e.g., \mathbf{p}), bold uppercase letters (e.g., \mathbf{P}), and calligraphic fonts (e.g., \mathcal{P}). Given the point cloud set \mathcal{P} , each point is represented as $\mathbf{p}_i \in \mathbb{R}^3$, ($i = 1, \dots, N$). The point cloud is organized into a feature matrix $\mathbf{P} \in \mathbb{R}^{N \times 3}$. Points cloud from \mathcal{P} can be encoded into tokens, yielding a token feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ with n tokens. $\mathbf{W} \in \mathbb{R}^{d \times r}$ is a linear projection matrix.

Overview Framework

The key challenge in continual 3D anomaly detection is constraining models to avoid catastrophic forgetting of past detection capabilities when facing new data. To this end, we propose a novel **Continual 3D Anomaly Detection (C3DAD)**. The overall framework is illustrated in Figure 2, comprising three main components: a Kernel Attention with random feature Layer (KAL), a Kernel Attention with learnable Advisor (KAA), and Reconstruction with Parameter Perturbation (RPP).

Kernel Attention with random feature Layer

Feature extraction is used to improve the performance of the downstream tasks. In the continual learning paradigm, it is essential to extract generalized features. These features aim to minimize the feature space differences across different tasks, mitigating catastrophic forgetting. In the scenario of 3D AD, point clouds suffer from poor structure and weak semantic information due to the organization as triplets in the form of $\mathbf{P} \in \mathbb{R}^{N \times 3}$. Hence, firstly, the local structure features are extracted for the generalized representation of the point clouds.

Specifically, \mathbf{P} is sampled into n centers group $\bar{\mathbf{P}}_{\text{center}}$ by the Furthest Point Sampling (FPS) (Charles et al. 2017). After sampling center points from the point cloud, the neighborhood point set of center point $\bar{\mathbf{p}}_i$ can be expressed as:

$$\mathcal{N}_r(\bar{\mathbf{p}}_i) = \{\mathbf{p}_j \in \mathbf{P} \mid \|\bar{\mathbf{p}}_i - \mathbf{p}_j\|_2 \leq r\}. \quad (1)$$

Due to varying point cloud scales across different classes, the radius r is adaptively adjusted using the following equation:

$$r = \frac{\eta}{|\bar{\mathbf{P}}_{\text{center}}|} \sum_{\bar{\mathbf{p}}_j \in \bar{\mathbf{P}}_{\text{center}}} \|\bar{\mathbf{p}}_i - \bar{\mathbf{p}}_j\|_2, \quad (2)$$

where η is the scaling factor of radius.

Secondly, the features should be mapped into a unified space. To tokenize the point cloud data, an encoder is employed. Without loss of generality, let \mathbf{q} , \mathbf{k} , and \mathbf{v} denote the query, key, and value vectors, respectively. The traditional self-attention mechanism utilized in encoders is:

$$\mathbf{o}_l = \sum_i^n \frac{\exp(\mathbf{q}_l^\top \mathbf{k}_i)}{\sum_j^n \exp(\mathbf{q}_l^\top \mathbf{k}_j)} \mathbf{v}_i. \quad (3)$$

Without considering the *softmax* activation function and scaling, the output is the optimum of $\min_{\mathbf{O}} \|\mathbf{O} - \mathbf{Q}\mathbf{K}^\top \mathbf{V}\|_F$, where \mathbf{K} and \mathbf{V} can be viewed as feature extraction and reconstruction matrix. The following kernel attention is proposed to capture nonlinear relationships among features:

$$\mathbf{o}_l = \sum_i^n \frac{\kappa(\mathbf{q}_l, \mathbf{k}_i)}{\sum_j^n \kappa(\mathbf{q}_l, \mathbf{k}_j)} \mathbf{v}_i, \quad (4)$$

where $\kappa(\mathbf{Q}, \mathbf{K})_{lj} = \kappa(\mathbf{q}_l, \mathbf{k}_j)$ is the kernel function. $\kappa(\mathbf{q}_l, \mathbf{k}_j)$ represents the inner product $\langle \phi(\mathbf{q}_l), \phi(\mathbf{k}_j) \rangle$, where $\phi(\cdot)$ is a mapping to the unified Hilbert space. If the space spanned by $\phi(\mathbf{K})$ is approximately unified, it can extract global generalized features in the continual learning paradigm. The output of the attention layer is:

$$\mathbf{o}_l = \sum_i^n \frac{\phi^\top(\mathbf{q}_l) \phi(\mathbf{k}_i)}{\sum_j^n \phi^\top(\mathbf{q}_l) \phi(\mathbf{k}_j)} \mathbf{v}_i. \quad (5)$$

However, the overall computational complexity is quadratic. To address this, it can be:

$$\mathbf{o}_l = \frac{(\sum_i^n \mathbf{v}_i \phi^\top(\mathbf{k}_i)) \phi(\mathbf{q}_l)}{\sum_j^n \phi^\top(\mathbf{k}_j) \phi(\mathbf{q}_l)}, \quad (6)$$

which is $\mathcal{O}(n)$ computational complexity in matrix form. The mapping ϕ can be defined by the random feature, e.g.,

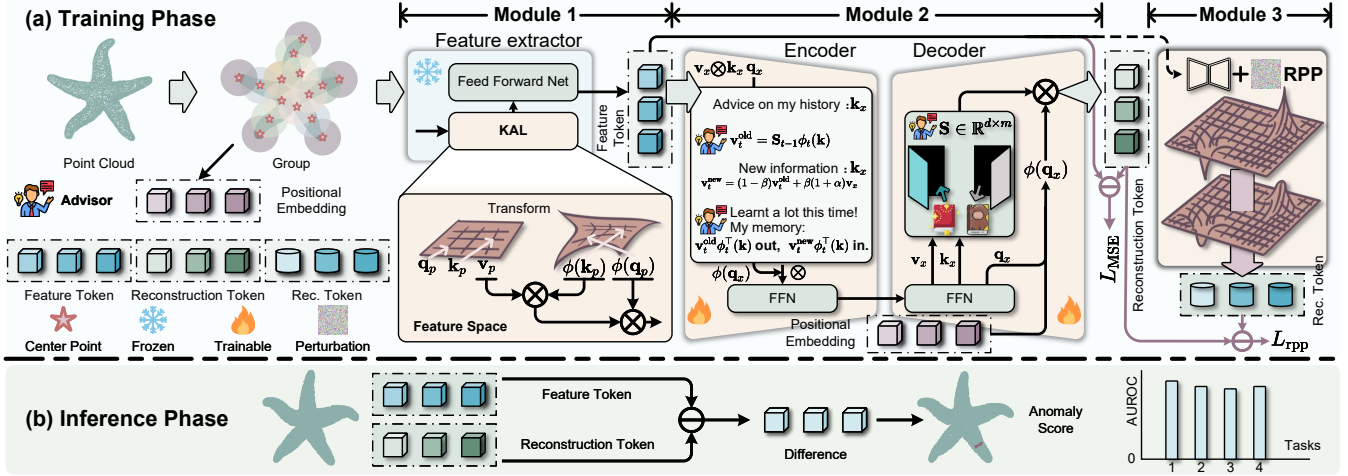


Figure 2: **The pipeline of C3D-AD.** The training point cloud data is aggregated into groups according to the center points. Feature tokens are generated by extracting features from both group centers and point groups, utilizing the Kernel Attention with random feature Layer (KAL) module in linear complexity. Then, feature tokens are input into the Encoder-Decoder, employing Kernel Attention with learnable Advisor (KAA) mechanism, which can memorize new class information and discard redundant information. Finally, the feature tokens are reconstructed again via Reconstruction with Parameter Perturbation (RPP) module, which can help the model review past samples and avoid catastrophic forgetting. Anomalies are detected according to the anomaly score by comparing the differences between feature tokens and reconstruction tokens.

Positive Random Feature (Choromanski et al. 2021):

$$\phi(\mathbf{x}) = \frac{e^{-\frac{\|\mathbf{x}\|^2}{2}} \left[e^{\mathbf{w}_1^\top \mathbf{x}}, \dots, e^{\mathbf{w}_m^\top \mathbf{x}} \right]}{\sqrt{m}}, \quad (7)$$

where the projection \mathbf{w}_i is sampled i.i.d. from $\mathbf{w}_i \sim \mathcal{N}(0, I_d)$. Usually, m is set to a small enough value.

In this way, KAL not only captures the local spatial context of the point cloud but also learn the generalized nonlinear structural information across inter-group point clouds.

Kernel Attention with learnable Advisor

After extracting generalized representations of point clouds locally and globally via the KAL, it is essential to introduce an advisor within the encoder-decoder architecture to mitigate catastrophic forgetting during continual learning. To address this, we introduce a novel continual learning attention mechanism with linear $\mathcal{O}(n)$ complexity, named Kernel Attention with learnable Advisor (KAA), which enables the model to learn new information efficiently while preserving previously acquired information.

Rewrite Eq. (6) by ignoring the denominator, and it becomes $\mathbf{o}_l = \mathbf{S}\phi(\mathbf{q}_l)$. $\mathbf{S} \in \mathbb{R}^{d \times m}$ is a continually learnable advisor. The following objective function (8) is proposed to train the advisor \mathbf{S} :

$$\min_{\mathbf{S}} L_{\text{kaa}}(\mathbf{S}) = \min_{\mathbf{S}} \frac{1}{2} \|\mathbf{S}\phi(\mathbf{k}) - \mathbf{v}\|^2 - \alpha \text{Tr}(\mathbf{v}^\top \mathbf{S}\phi(\mathbf{k})), \quad (8)$$

where the advisor \mathbf{S} guides the key $\phi(\mathbf{k})$ close to the value \mathbf{v} and aligns their directions. Getting the derivative of the function L_{kaa} w.r.t. \mathbf{S} and setting it to zero, the update gradient can be derived as $\nabla_{\mathbf{S}} L_{\text{kaa}} = \mathbf{S}\phi(\mathbf{k})\phi^\top(\mathbf{k}) - (1 + \alpha)\phi(\mathbf{k})\mathbf{v}^\top$.

Given the learning rate β , the update rule is:

$$\begin{aligned} \mathbf{S}_t &= \mathbf{S}_{t-1} - \beta \nabla_{\mathbf{S}} L_{\text{kaa}} \\ &= \mathbf{S}_{t-1} - \beta (\mathbf{S}_{t-1} \phi(\mathbf{k}) \phi^\top(\mathbf{k}) - (1 + \alpha) \phi(\mathbf{k}) \mathbf{v}^\top). \end{aligned} \quad (9)$$

Let $\mathbf{v}_t^{\text{old}} = \mathbf{S}_{t-1} \phi_t(\mathbf{k})$ and $\mathbf{v}_t^{\text{new}} = (1 - \beta) \mathbf{v}_t^{\text{old}} + \beta (1 + \alpha) \mathbf{v}_t$, then Eq. (9) becomes:

$$\mathbf{S}_t = \mathbf{S}_{t-1} - \mathbf{v}_t^{\text{old}} \phi_t^\top(\mathbf{k}) + \mathbf{v}_t^{\text{new}} \phi_t^\top(\mathbf{k}), \quad (10)$$

where $-\mathbf{v}_t^{\text{old}} \phi_t^\top(\mathbf{k})$ represents that reduces redundant information from previous tasks and $\mathbf{v}_t^{\text{new}} \phi_t^\top(\mathbf{k})$ is to learn new information. Hence, the output of attention in the t -th task is:

$$\mathbf{O}_t = \phi(\mathbf{Q}_t) \mathbf{S}_t^\top, \quad (11)$$

where l -th row of $\phi(\mathbf{Q}_t)$ is $\phi^\top(\mathbf{q}_l)$.

Storing historical point cloud information in \mathbf{S} preserves KAA's continual learning capability. It eliminates the need to maintain a set of past samples. Furthermore, due to its linear complexity, the proposed method is resource-efficient.

Reconstruction with Parameter Perturbation

KAA is to learn new information while reducing previous one for the Encoder-Decoder. In this section, the Reconstruction with Parameter Perturbation (RPP) mechanism is proposed to reconstruct data from the view of the future. To ensure that the model converges globally to a hypothesis space that is optimal across all tasks, the gradient of the model's parameters must be continually constrained to optimize within the intersection of the optimal hypotheses for each task. This strategy aims to prevent catastrophic forgetting. According to the objective, the model should maintain a high similarity for the same batch of data at time t and

time $t + \Delta t$, $\Delta t > 0$. Consequently, this is formulated as the following minimization of a loss:

$$L_{\text{rpp}} = \|h(\theta_t, x) - h(\theta_{t+\Delta t}, x)\|_2^2. \quad (12)$$

where $h(\theta_t, \cdot)$ is the hypothesis at time t . However, it is impossible to obtain the hypothesis $h(\theta_{t+\Delta t}, \cdot)$ at time t . Hence, the state after Δt must be predicted based on the current model weights. To this end, we approximate the hypothesis after Δt as $h(\theta_t + \delta)$, where the initial perturbation δ is sampled from a normal distribution and constrained by $\|\delta\|_2 \leq \epsilon$. Within the parameter space, the perturbation is searched to induce the worst-case deviation upon addition, thereby simulating the most adverse future scenario. Motivated by this, the optimization problem is proposed:

$$\begin{aligned} L_{\text{rpp}}(\theta_t, \mathbf{P}_t) = \max_{\delta} \|h(\theta_t, \mathbf{P}_t) - h(\theta_t + \delta, \mathbf{P}_t)\|_2^2 \\ \text{s.t. } \|\delta\|_2 \leq \epsilon. \end{aligned} \quad (13)$$

To find a local optimum of this objective function, gradient ascent is employed.

Generalization Error Bound of RPP Loss To assess the generalization error incurred by the objective function when using limited point cloud data to converge the model to the concept set, the generalization bound is derived for L_{rpp} .

Theorem 1. *Let $\mathcal{G} = \{g : x \mapsto \|h(\theta, x) - h(\theta + \delta, x)\|_2^2 \mid \|\delta\|_2 \leq \epsilon\}$ be the class of function induced by bounded perturbations δ . For all $g \in \mathcal{G}$ and any input x , with probability at least $1 - \xi$ over the random draw of a training sample of size N from the underlying distribution \mathcal{D} , the following holds:*

$$\hat{L}_{\text{rpp}}(\theta) \leq L_{\text{rpp}}(\theta) + 2\epsilon^2 L_{\theta}^2 \mathcal{R}(\psi) + 3\sqrt{\frac{\log(2/\xi)}{2N}}. \quad (14)$$

According to Theorem 1, the perturbation constraint ϵ should not be excessively large to avoid a significant increase in the generalization error. More details can be found in the *Supplemental Materials*.

Experiments

Experimental setting

Datasets. The **Real3D-AD** (Liu et al. 2023) is a benchmark for 3D AD, comprising 1,254 large-scale, high-resolution samples from 12 object categories. The training set for each category consists of only 4 normal samples, and the test set contains both normal samples and various anomalies. The **Anomaly-ShapeNet** (Li et al. 2024) is a large-scale synthetic dataset for AD. It comprises 1,600 samples distributed across 40 object categories, posing a significant challenge due to its high inter-class diversity. **MulSen-AD** (Li et al. 2025a) is the high-resolution multi-sensor anomaly detection data set. It consists of 2,035 samples from 15 industrial object categories, which are split into a training set of 1,391 normal samples and a test set comprising 150 normal and 494 anomalous samples.

Comparing baselines. We selected classical 3D anomaly detection methods: BTF (Horwitz and Hoshen 2023), M3DM (Wang et al. 2023), Patchcore (Roth et al. 2022), CPMF (Cao, Xu, and Shen 2024), Reg3D-AD (Liu et al. 2023), IMRNet (Li et al. 2024), R3D-AD (Zhou et al. 2025), MC3D-AD (Cheng et al. 2025), PLANE (Wang et al. 2025), and PO3AD (Ye et al. 2024) to demonstrate that C3D-AD is effective in anomaly detection. Moreover, we modified three baseline methods adapted for continual learning for fair comparison: Continual-PatchCore, Continual-MC3D-AD, and Continual-Reg3D-AD. Complete implementation details are provided in the *Supplementary Materials*.

Continual Learning Setting To estimate the models’ performance in class-incremental 3D Anomaly Detection, we partitioned each dataset into a series of sequential tasks. For the Real3D-AD dataset, we divide the training set into 4 disjoint tasks based on categories. The test sets are constructed cumulatively, i.e., $\mathcal{P}_{\text{test}}^1 \subset \dots \subset \mathcal{P}_{\text{test}}^4$. Following this, Anomaly-ShapeNet and MulSen-AD datasets are divided into 4 and 3 tasks, respectively. To evaluate the performance of the model, we adopt metrics at the object level. For object-level AD, the AUROC (\uparrow) is employed. To ensure a fair comparison across all categories, the mean AUROC and average ranking of each method (\downarrow) are reported. Code would be available upon acceptance of the paper.

Implementation Details. PointMAE (Pang et al. 2022) with the KAL is pre-trained on ModelNet408K (Wu et al. 2015) for feature extraction. The α and β of KAA are both set to 0.7, and m is set to 10. The sample scaling factor η is 10 to cover the whole point cloud. The ϵ is scanned in the range $[0.01, 10]$, determining the generalization error. The AdamW optimizer is employed in the training process with initial learning rate 0.0001 and rate 0.00001 after 800 epochs. The number of stacked encoder-decoder blocks is set to 4. Our experiments were conducted on a machine with PyTorch 1.13.0, CUDA 11.7, and an NVIDIA A100-PCIE-40GB GPU.

Performance on Continual Anomaly Detection

The experimental results of C3D-AD and other methods in the continual learning paradigm are shown in Table 1. C3D-AD has demonstrated state-of-the-art (SOTA) performance in 3D Anomaly Detection across various tasks and datasets. Unlike memory-bank-based methods, e.g., continual-Reg3D-AD and continual-PatchCore, C3D-AD achieves superiority by unifying the feature space via KAL, while minimizing reconstruction error via RPP. Hence, C3D-AD outperform these two method by 14.3%, 31.2%, and 5.4% on three dataset. Furthermore, while the unified model MC3D-AD implicitly maintains previous information, C3D-AD performs better. The gaps, which are 3.9%, 3.4%, and 3.0%, are attributed to the lack of mechanisms like KAA encoding past information into the model’s parameters and the RPP module for information rehearsal. The experiments also demonstrate that C3D-AD can be applied in class-incremental AD.

	<i>Real3D-AD</i>				<i>Anomaly-ShapeNet</i>				<i>MuSen-AD</i>		
task_id	1	2	3	4	1	2	3	4	1	2	3
Continual-Reg3D-AD	0.598	0.527	0.505	0.521	0.522	0.544	0.501	0.519	0.602	0.599	0.580
Continual-PatchCore (FPFH)	0.566	0.527	0.500	0.518	0.567	0.541	0.510	0.494	0.491	0.549	0.524
Continual-MC3D-AD	0.767	0.650	0.629	0.625	0.829	0.803	0.833	0.797	0.623	0.626	0.604
C3D-AD	0.805	0.658	0.666	0.664	0.854	0.848	0.862	0.831	0.698	0.650	0.634

Table 1: The mean O-AUROC (\uparrow) performance of different methods across multiple tasks of datasets. The best results are **bold**.

Method	cap0	cap3	helmet3	cup0	bow14	vase3	headset1	eraser0	vase8	cap4	vase2	vase4	helmet0	bucket1
BTF(Raw)	0.668	0.527	0.526	0.403	0.664	0.717	0.515	0.525	0.424	0.468	0.410	0.425	0.553	0.321
BTF(FPFH)	0.618	0.522	0.444	0.586	0.609	0.699	0.490	0.719	0.668	0.520	0.546	0.510	0.571	0.633
M3DM	0.557	0.423	0.374	0.539	0.464	0.439	0.617	0.627	0.663	0.777	0.737	0.476	0.526	0.501
Patchcore(FPFH)	0.580	0.453	0.404	0.600	0.494	0.449	0.637	0.657	0.662	0.757	0.721	0.506	0.546	0.551
Patchcore(PointMAE)	0.589	0.476	0.424	0.610	0.501	0.460	0.627	0.677	0.663	0.727	0.741	0.516	0.556	0.561
CPMF	0.601	0.551	0.520	0.497	0.683	0.582	0.458	0.689	0.529	0.553	0.582	0.514	0.555	0.601
Reg3D-AD	0.693	0.725	0.367	0.510	0.663	0.650	0.610	0.343	0.620	0.643	0.605	0.500	0.600	0.752
IMRNet	0.737	0.775	0.573	0.643	0.676	0.700	0.676	0.548	0.630	0.652	0.614	0.524	0.597	0.771
R3D-AD	0.822	0.730	0.707	0.776	0.744	0.742	0.795	0.890	0.721	0.681	0.752	0.630	0.757	0.756
MC3D-AD	0.793	0.701	0.979	0.743	<u>0.911</u>	0.761	<u>0.886</u>	0.776	0.670	<u>0.835</u>	<u>0.929</u>	0.876	0.672	0.784
PLANE	0.944	0.954	0.721	<i>0.805</i>	<i>0.893</i>	0.782	0.776	1.000	0.964	0.730	0.971	0.773	0.704	0.968
PO3AD	<u>0.877</u>	<u>0.859</u>	<i>0.754</i>	<u>0.871</u>	0.981	0.821	0.923	<u>0.995</u>	<i>0.739</i>	<i>0.792</i>	<u>0.952</u>	0.675	<u>0.762</u>	0.787
C3D-AD	0.678	<i>0.800</i>	<u>0.900</u>	0.929	0.626	<u>0.791</u>	0.752	0.843	<u>0.773</u>	0.884	0.919	<u>0.836</u>	0.870	<u>0.863</u>

Method	bottle3	vase0	bottle0	tap1	bow10	bucket0	vase5	vase1	vase9	ashtray0	bottle1	tap0	phone	cup1
BTF(Raw)	0.568	0.531	0.597	0.573	0.564	0.617	0.585	0.549	0.564	0.578	0.510	0.525	0.563	0.521
BTF(FPFH)	0.322	0.342	0.344	0.546	0.509	0.401	0.409	0.219	0.268	0.420	0.546	0.560	0.671	0.610
M3DM	0.541	0.423	0.574	0.739	0.634	0.309	0.317	0.427	0.663	0.577	0.637	<i>0.754</i>	0.357	0.556
Patchcore(FPFH)	0.572	0.455	0.604	0.766	0.504	0.469	0.417	0.423	0.660	0.587	0.667	0.753	0.388	0.586
Patchcore(PointMAE)	0.650	0.447	0.513	0.538	0.523	0.593	0.579	0.552	0.629	0.591	0.601	0.458	0.488	0.556
CPMF	0.405	0.451	0.520	0.697	0.783	0.482	0.618	0.345	0.609	0.353	0.482	0.359	0.509	0.499
Reg3D-AD	0.525	0.533	0.486	0.641	0.671	0.610	0.520	0.702	0.594	0.597	0.695	0.676	0.414	0.538
IMRNet	0.640	0.533	0.552	0.696	0.681	0.580	0.676	<i>0.757</i>	0.594	0.671	0.700	0.676	0.755	0.757
R3D-AD	0.781	0.788	0.733	<u>0.900</u>	0.819	0.683	0.757	0.729	0.718	0.833	0.737	0.736	0.762	0.757
MC3D-AD	0.756	0.821	0.795	0.970	<u>0.930</u>	<u>0.898</u>	0.976	<u>0.857</u>	<i>0.736</i>	<u>0.962</u>	0.709	<u>0.945</u>	<i>0.919</i>	<u>0.952</u>
PLANE	0.994	0.896	<i>0.843</i>	0.652	0.963	0.981	0.690	<i>0.771</i>	0.592	<i>0.905</i>	<u>0.814</u>	0.467	1.000	0.705
PO3AD	<u>0.926</u>	<i>0.858</i>	<u>0.900</u>	0.681	<i>0.922</i>	<i>0.853</i>	<u>0.852</u>	0.742	<u>0.830</u>	1.000	0.933	0.745	0.776	<i>0.833</i>
C3D-AD	<i>0.857</i>	<u>0.888</u>	0.962	<i>0.844</i>	0.919	0.822	<i>0.795</i>	0.938	0.897	0.752	<i>0.740</i>	0.955	<u>0.938</u>	0.967

Method	vase7	helmet2	cap5	shelf0	bow15	bow13	helmet1	bow11	headset0	bag0	bow12	jar	Mean	A. R.
BTF(Raw)	0.448	0.602	0.373	0.164	0.417	0.385	0.349	0.264	0.378	0.410	0.525	0.420	0.493	10.68
BTF(FPFH)	0.518	0.542	0.586	0.609	0.699	0.490	0.719	0.668	0.520	0.546	0.510	0.424	0.528	9.98
M3DM	0.657	0.623	0.639	0.564	0.409	0.617	0.427	0.663	0.577	0.537	0.684	0.441	0.552	9.73
Patchcore(FPFH)	0.693	0.425	<i>0.790</i>	0.494	0.558	0.537	0.484	0.639	0.583	0.571	0.615	0.472	0.568	9.25
Patchcore(PointMAE)	0.650	0.447	0.538	0.523	0.593	0.579	0.552	0.629	0.591	0.601	0.458	0.483	0.562	9.35
CPMF	0.397	0.462	0.697	0.685	0.685	0.658	0.589	0.639	0.643	0.643	0.625	0.610	0.559	9.25
Reg3D-AD	0.462	0.614	0.467	0.688	0.593	0.348	0.381	0.525	0.537	0.706	0.490	0.592	0.572	9.30
IMRNet	0.635	0.641	0.652	0.603	0.710	0.599	0.600	0.702	0.720	0.660	0.685	0.780	0.661	6.73
R3D-AD	0.771	0.633	0.670	0.696	0.656	0.767	0.720	0.778	0.738	0.720	0.741	0.838	0.749	4.63
MC3D-AD	<u>0.938</u>	0.609	0.761	0.841	0.754	0.885	1.000	0.978	0.862	0.805	0.719	<u>0.971</u>	<u>0.842</u>	<i>3.05</i>
PLANE	<u>0.938</u>	1.000	0.870	<i>0.759</i>	<i>0.800</i>	0.706	0.543	<u>0.907</u>	0.782	0.914	0.956	1.000	0.836	3.18
PO3AD	0.966	<u>0.869</u>	0.670	0.573	<u>0.849</u>	<u>0.881</u>	<i>0.961</i>	<i>0.829</i>	<i>0.808</i>	<u>0.833</u>	<u>0.833</u>	0.866	0.839	<u>2.88</u>
C3D-AD	0.900	<i>0.745</i>	<u>0.811</u>	<u>0.765</u>	0.944	<i>0.793</i>	<u>0.971</u>	0.711	<u>0.827</u>	<u>0.905</u>	<i>0.822</i>	<i>0.924</i>	0.846	2.78

Table 2: The O-AUROC (\uparrow) performance of different methods across 40 categories of Anomaly-ShapeNet. The best, second-best, and third-best are **bold**, underline, and *italics*, respectively. **A. R.** represents the average ranking of each method.

Performance on Multi-class Anomaly Detection

The capability of C3D-AD for continual 3D AD indicates that it has the intrinsic ability to learn from various data

classes. To access the baseline of C3D-AD, multi-class anomaly detection is conducted, which is not on a continual learning paradigm. The experimental results are shown

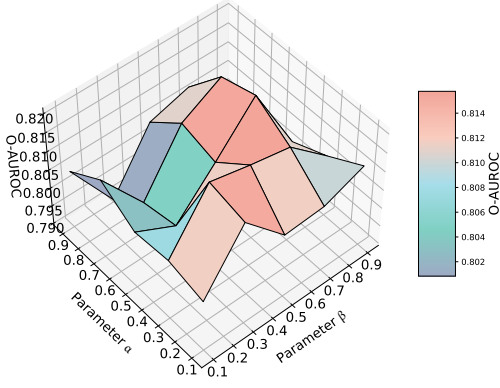


Figure 3: Sensitivity of α and β on Anomaly-ShapeNet.

in Table 2. Compared to other methods, C3D-AD has the superiority on average AUROC, improving 0.4% than MC3D-AD. In addition, C3D-AD achieves the top average ranking in 40 classes. This demonstrates its satisfactory performance in multi-class anomaly detection scenarios.

Ablation Study

To demonstrate the effectiveness of our proposed modules, ablation studies are conducted on KAL, KAA, and RPP. The experimental results, as presented in Table 3, show that the model achieves its best performance on continual 3D AD only when all three components are integrated. Based on the results across the datasets, KAL and RPP show a more significant performance. This suggests that extracting generalized features and maintaining feature consistency are key to enhancing the continual 3D AD. Furthermore, the information from the advisor is also significant for the model to learn new data and avoid catastrophic forgetting continually.

KAL	KAA	RPP	R.	A.	M.
×	✓	✓	0.534	0.791	0.612
✓	×	✓	0.565	0.814	0.607
✓	✓	×	0.529	0.798	0.557
✓	✓	✓	0.664	0.831	0.634

Table 3: Results of ablation study on Real3D-AD (R.), Anomaly-ShapeNet (A.), and MulSen-AD (M.).

$\epsilon =$	10^{-2}	10^{-1}	10^0	10^1	10^2	10^3
AUROC	0.801	0.817	0.787	0.809	0.780	0.769

Table 4: Sensitivity analysis of ϵ on Anomaly-ShapeNet.

Sensitivity of Parameters

The sensitivity experiments evaluate the value of hyperparameters. In KAA, α is to align the direction of information from the advisor and the value vector, and β is to leverage

$m =$	10^1	10^2	10^3	5×10^3
AUROC	0.802	0.828	0.842	0.819
Inf. T. (s)	0.281	0.311	0.393	1.109
Memory (GB)	4.431	4.650	5.680	9.400

Table 5: Analysis of m on Anomaly-ShapeNet. The inference time (Inf. T.) and the GPU Memory usage are reported.

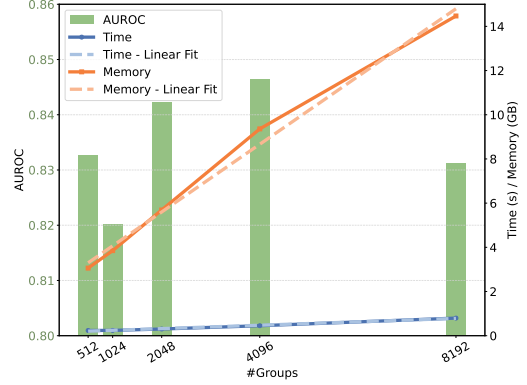


Figure 4: Inference time (s) and memory usage (GB) vs. #Groups for C3D-AD on Anomaly-ShapeNet.

the previous and new information. As shown in Figure 3, α and β are set to 0.7, which can achieve satisfactory performance. In addition, ϵ is an important hyperparameter of RPP. According to Theorem 1, ϵ should not be too large, which is demonstrated as shown in Table 4. The hyperparameter m in KAL and KAA is usually a small value. As shown in Table 5, the computational complexity w.r.t. m is $\mathcal{O}(m)$. However, larger m , which is the attribute of the advisor, may not lead to better performance. To leverage the performance and efficiency, m of KAL and KAA can be set from 10^1 to 10^3 .

Empirical Computational Complexity Analysis

Due to the linear computational complexity of Kernel Attention in KAL and KAA, C3D-AD is efficient in anomaly detection. Figure 4 shows that the inference time and GPU peak memory increase linearly with respect to the group number, which is relevant to the number of tokens. The performance can be satisfactory by setting the number to 4096. Hence, C3D-AD is tailored for industrial continual 3D AD due to its $\mathcal{O}(n)$ complexity and performance.

Conclusions

The continuous emergence of new object categories poses a significant challenge for 3D Anomaly Detection. To address this, in this study, we propose a continual learning framework named C3D-AD. Firstly, we introduce a Kernel Attention with random feature Layer to extract generalized features. Then, to reconstruct feature tokens while avoiding catastrophic forgetting, a Kernel Attention with learnable Advisor module is designed within the encoder-decoder to learn new information while discarding redundant one. Furthermore, to maintain representation consis-

tency across tasks, the tokens are reconstructed again using Reconstruction with Parameter Perturbation, which aligns the model’s current and future outputs. Experiments on benchmark datasets demonstrate our method’s superiority in continual 3D AD, achieving state-of-the-art performance with satisfactory efficiency. For future work, further research is needed to explore how to effectively constrain the advisor to achieve better performance in continual 3D AD.

References

- Barusco, M.; D’Antoni, L.; Pezze, D. D.; Borsatti, F.; and Susto, G. A. 2025. Memory Efficient Continual Learning for Edge-Based Visual Anomaly Detection. *arXiv:2503.02691*.
- Batra, H.; and Clark, R. 2024. EVCL: Elastic Variational Continual Learning with Weight Consolidation. *arXiv:2406.15972*.
- Bergmann, P.; and Sattlegger, D. 2023. Anomaly Detection in 3D Point Clouds Using Deep Geometric Descriptors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2613–2623.
- Bonicelli, L.; Boschini, M.; Porrello, A.; Spampinato, C.; and Calderara, S. 2022. On the Effectiveness of Lipschitz-Driven Rehearsal in Continual Learning. In Oh, A. H.; Agarwal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.
- Cao, Y.; Xu, X.; and Shen, W. 2024. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition(PR)*, 156: 110761.
- Charles, R. Q.; Su, H.; Kaichun, M.; and Guibas, L. J. 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 77–85.
- Chen, R.; Xie, G.; Liu, J.; Wang, J.; Luo, Z.; Wang, J.; and Zheng, F. 2023. EasyNet: An Easy Network for 3D Industrial Anomaly Detection. *arXiv:2307.13925*.
- Cheng, J.; Gao, C.; Zhou, J.; Wen, J.; Dai, T.; and Wang, J. 2025. MC3D-AD: A Unified Geometry-aware Reconstruction Model for Multi-category 3D Anomaly Detection. *arXiv:2505.01969*.
- Choromanski, K. M.; Likhoshesterov, V.; Dohan, D.; Song, X.; Gane, A.; Sarlos, T.; Hawkins, P.; Davis, J. Q.; Mohiuddin, A.; Kaiser, L.; Belanger, D. B.; Colwell, L. J.; and Weller, A. 2021. Rethinking Attention with Performers. In *International Conference on Learning Representations*.
- Eskandar, M.; Imtiaz, T.; Hill, D.; Wang, Z.; and Dy, J. 2025. STAR: Stability-Inducing Weight Perturbation for Continual Learning. In *The Thirteenth International Conference on Learning Representations*.
- Gu, Z.; Zhang, J.; Liu, L.; Chen, X.; Peng, J.; Gan, Z.; Jiang, G.; Shu, A.; Wang, Y.; and Ma, L. 2024. Rethinking Reverse Distillation for Multi-Modal Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8): 8445–8453.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 6840–6851. Curran Associates, Inc.
- Horwitz, E.; and Hoshen, Y. 2023. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2968–2977.
- Hu, L.; Gan, Z.; Deng, L.; Liang, J.; Liang, L.; Huang, S.; and Chen, T. 2025. ReplayCAD: Generative Diffusion Replay for Continual Anomaly Detection. *arXiv preprint arXiv:2505.06603*.
- Kingma, D. P.; and Welling, M. 2022. Auto-Encoding Variational Bayes. *arXiv:1312.6114*.
- Kruse, M.; Rudolph, M.; Woiwode, D.; and Rosenhahn, B. 2024. SplatPose & Detect: Pose-Agnostic 3D Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 3950–3960.
- Li, W.; Xu, X.; Gu, Y.; Zheng, B.; Gao, S.; and Wu, Y. 2024. Towards Scalable 3D Anomaly Detection and Localization: A Benchmark via 3D Anomaly Synthesis and A Self-Supervised Learning Network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22207–22216.
- Li, W.; Zhan, J.; Wang, J.; Xia, B.; Gao, B.-B.; Liu, J.; Wang, C.; and Zheng, F. 2022. Towards Continual Adaptation in Industrial Anomaly Detection. In *Proceedings of the 30th ACM International Conference on Multimedia*, 2871–2880. Association for Computing Machinery.
- Li, W.; Zheng, B.; Xu, X.; Gan, J.; Lu, F.; Li, X.; Ni, N.; Tian, Z.; Huang, X.; Gao, S.; and Wu, Y. 2025a. Multi-Sensor Object Anomaly Detection: Unifying Appearance, Geometry, and Internal Properties. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, 9984–9993.
- Li, X.; Tan, X.; Chen, Z.; Zhang, Z.; Zhang, R.; Guo, R.; Jiang, G.; Chen, Y.; Qu, Y.; Ma, L.; and Xie, Y. 2025b. One-for-More: Continual Diffusion Model for Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4766–4775.
- Liang, H.; Xie, G.; Hou, C.; Wang, B.; Gao, C.; and Wang, J. 2025a. Look Inside for More: Internal Spatial Modality Perception for 3D Anomaly Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(5): 5146–5154.
- Liang, H.; Zhang, J.; Dai, T.; Shen, L.; Wang, J.; and Gao, C. 2025b. Taming Anomalies with Down-Up Sampling Networks: Group Center Preserving Reconstruction for 3D Anomaly Detection. *arXiv:2507.03903*.
- Liu, J.; Wu, K.; Nie, Q.; Chen, Y.; Gao, B.-B.; Liu, Y.; Wang, J.; Wang, C.; and Zheng, F. 2024a. Unsupervised Continual Anomaly Detection with Contrastively-Learned Prompt. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(4): 3639–3647.
- Liu, J.; Xie, G.; Chen, R.; Li, X.; Wang, J.; Liu, Y.; Wang, C.; and Zheng, F. 2023. Real3D-AD: A Dataset of Point Cloud Anomaly Detection. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in*

- Neural Information Processing Systems*, volume 36, 30402–30415.
- Liu, J.; Xie, G.; Wang, J.; Li, S.; Wang, C.; Zheng, F.; and Jin, Y. 2024b. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1): 104–135.
- Mallya, A.; Davis, D.; and Lazebnik, S. 2018. Piggyback: Adapting a Single Network to Multiple Tasks by Learning to Mask Weights. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Mallya, A.; and Lazebnik, S. 2018. PackNet: Adding Multiple Tasks to a Single Network by Iterative Pruning. arXiv:1711.05769.
- Masuda, M.; Hachiuma, R.; Fujii, R.; Saito, H.; and Sekikawa, Y. 2021. Toward Unsupervised 3d Point Cloud Anomaly Detection Using Variational Autoencoder. In *2021 IEEE International Conference on Image Processing (ICIP)*, 3118–3122.
- Pang, Y.; Wang, W.; Tay, F. E.; Liu, W.; Tian, Y.; and Yuan, L. 2022. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, 604–621. Springer.
- Qin, J.; Gu, C.; Yu, J.; and Zhang, C. 2023. Teacher–student network for 3D point cloud anomaly detection with few normal samples. *Expert Systems with Applications*, 228: 120371.
- Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. iCaRL: Incremental Classifier and Representation Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Riemer, M.; Klinger, T.; Bouneffouf, D.; and Franceschini, M. 2019. Scalable recollections for continual lifelong learning. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI Press.
- Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; and Gehler, P. 2022. Towards Total Recall in Industrial Anomaly Detection. arXiv:2106.08265.
- Rudolph, M.; Wehrbein, T.; Rosenhahn, B.; and Wandt, B. 2023. Asymmetric Student-Teacher Networks for Industrial Anomaly Detection. In *Winter Conference on Applications of Computer Vision (WACV)*.
- Tang, J.; Lu, H.; Xu, X.; Wu, R.; Hu, S.; Zhang, T.; Cheng, T. W.; Ge, M.; Chen, Y.-C.; and Tsung, F. 2025. An Incremental Unified Framework for Small Defect Inspection. In *Computer Vision – ECCV 2024*, 307–324.
- Tu, Y.; Zhang, B.; Liu, L.; Li, Y.; Zhang, J.; Wang, Y.; Wang, C.; and Zhao, C. 2025. Self-supervised Feature Adaptation for 3D Industrial Anomaly Detection. In *Computer Vision – ECCV 2024*, 75–91.
- Wang, J.; Xu, H.; Chen, X.; Xu, H.; Huang, Y.; Ding, X.; and Tu, X. 2025. Exploiting Point-Language Models with Dual-Prompts for 3D Anomaly Detection. arXiv:2502.11307.
- Wang, Y.; Peng, J.; Zhang, J.; Yi, R.; Wang, Y.; and Wang, C. 2023. Multimodal Industrial Anomaly Detection via Hybrid Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8032–8041.
- Wang, Z.; Jian, T.; Chowdhury, K.; Wang, Y.; Dy, J.; and Ioannidis, S. 2020. Learn-Prune-Share for Lifelong Learning. In *2020 IEEE International Conference on Data Mining (ICDM)*, 641–650.
- Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; and Xiao, J. 2015. 3D ShapeNets: A Deep Representation for Volumetric Shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ye, J.; Zhao, W.; Yang, X.; Cheng, G.; and Huang, K. 2024. PO3AD: Predicting Point Offsets toward Better 3D Point Cloud Anomaly Detection. arXiv:2412.12617.
- Zhou, Z.; Wang, L.; Fang, N.; Wang, Z.; Qiu, L.; and Zhang, S. 2025. R3D-AD: Reconstruction via Diffusion for 3D Anomaly Detection. In *Computer Vision – ECCV 2024*, 91–107.