GMETRIC-3D: GEOMETRIC METRIC ENCODING BASED ON TOPOLOGICAL CONSISTENCY FOR 3D INDUSTRIAL DEFECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

The 3D industrial surface defect detection is a challenging task due to the small scale, irregular shape, and lack of stable semantic patterns of the defects. However, the existing 3D defect detection methods mainly rely on semantic and appearance features and fail to effectively capture these subtle geometric variations. To address this issue, we propose the geometric metric encoding based on topological consistency (GMetric-3D) that explicitly models the local geometric structure in defects. Specifically, we first introduce the geometric metric encoding to capture local structural patterns by combining metrics such as relative positions, normal vector variations, and curvature information within the point neighborhood. To further enhance the distinguishability of defect features, we present the topological consistency learning, which retains topological structure in normal regions and expands feature differences in defect regions through topological smoothing, separation, and boundary perception constraints. Extensive experiments on the high-precision 3D point cloud dataset CPS3D-Det and the large-scale scene dataset KITTI demonstrate that our GMetric-3D achieves state-of-the-art performance compared with existing 3D detection methods. The experimental results confirm that modeling geometric structure information is crucial for reliable and accurate 3D industrial defect detection. The source code will be released.

1 INTRODUCTION

The 3D defect detection requires the ability to identify defects in the complex structures of industrial products. These defects typically exhibit unique attributes such as small size, complex structure, and irregular distribution. The traditional methods primarily rely on 2D images for analysis, while 3D point cloud data provides more abundant geometric structure information. Nevertheless, recent 3D point cloud-based defect detection methods mainly focus on semantic features and appearance patterns, which have significant limitations in industrial defect detection scenarios. Industrial defects generally lack stable semantic categories, and their essence is manifested as abnormal changes in local geometric structures. Furthermore, the structural defects are often difficult to effectively capture by traditional feature learning paradigms.

Some studies (Zhao et al., 2023; Cheng et al., 2026; Li et al., 2025a) adapt to defective objects by enhancing feature representations. PFENet (Li et al., 2025a) merges low-level spatial features with high-level semantic features to expand the receptive field and capture multi-scale features. Despite these advances, many methods still rely on intuitive geometric structure design and lack the analysis of the intrinsic representation of defect objects. Some existing methods mitigate the impact of noise by introducing virtual points (Zhu et al., 2022; Wu et al., 2023) or more realistic geometric representations (Cao et al., 2024; Asad et al., 2025; Tseng et al., 2023). Due to the complementarity between 2D and 3D modal features, CPMF (Cao et al., 2024), 2M3DF(Asad et al., 2025) and, CrossDTR (Tseng et al., 2023) enhance the 3D spatial geometric representations by introducing multi-view 2D images. In addition, there are studies that utilize semi-supervised learning (Liu et al., 2023; Wang et al., 2024a; Yang et al., 2025) to improve the detection performance for unknown defects. However, these methods primarily rely on object-level and semantic-level representations, lacking the explicit modeling of topological consistency and structural mutations of defects, thereby limiting the generalization ability of the model in complex industrial scenarios.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

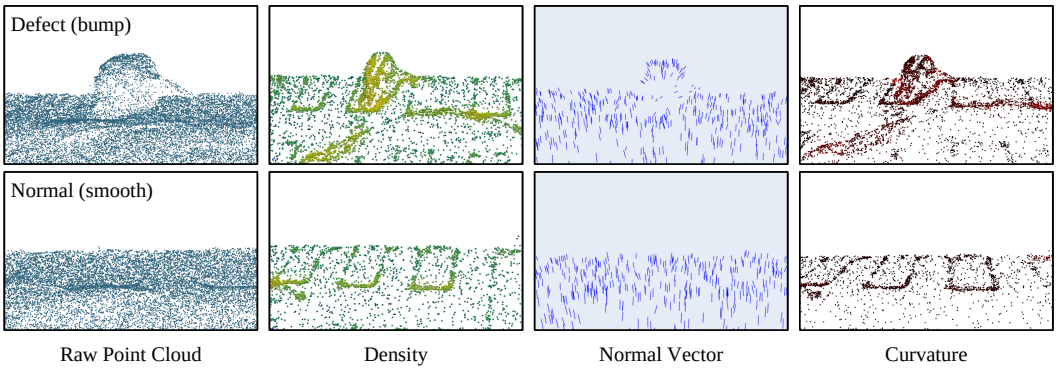


Figure 1: Illustrations of the local geometric features in the normal area and the defective area. The defects are represented as abnormal deviations from the local structural patterns. We visualize point cloud density, normal vector information, and curvature variations, respectively.

To address this issue, this paper presents the geometric metric encoding based on topological consistency for 3D industrial defect detection, focusing on the geometric mutation features and topological consistency of the defects as shown in Fig. 1. The goal of our detection task is similar to that of recent methods (Yu et al., 2026; Yu & Zhao, 2026), which treat defects as general objects to locate and regress the 3D bounding boxes. This is completely different from 3D anomaly detection. The proposed GMetric-3D focuses on the intrinsic attributes of defect occurrence in real-world industrial production scenarios. Specifically, we first introduce the geometric metric encoding, which constructs local neighborhood relationships between points and combines the metrics such as relative positions, normal vector variations, and curvature information of 3D spatial features to extract discriminative local structural features. Furthermore, we propose the topological consistency learning. It constrains the topological consistency of normal regions in the feature space, and on the other hand, it enhances the structural differences between defect regions, thereby effectively improving the discriminative ability of the feature space. In summary, the major contributions of this paper are listed as follows:

- We propose the geometric metric encoding based on topological consistency (GMetric-3D) for 3D industrial defect detection, focusing on the abnormal deviations in the local structural patterns of defects, explicitly capturing and utilizing local geometric relationships.
- We introduce the geometric metric encoding that combines the relative positions, normal vector variations, and curvature information of 3D spatial features to construct the local neighborhood relationships of point pairs and further enhances the discriminative ability of the feature space through topological consistency learning.
- We demonstrate the effectiveness of the proposed GMetric-3D through extensive experiments. Specifically, we conducted evaluations on the high-precision 3D point cloud dataset CPS3D-Det and the large-scale scene dataset KITTI.

2 Related Work

3D object detection in point clouds. 3D object detection based on LiDAR point clouds has received extensive research in recent years. The mainstream 3D object detection methods have three baselines: voxel-based methods (Wang et al., 2024b; Liu et al., 2025; Prasanth et al., 2026), point-based methods (Zhao et al., 2021; Ye et al., 2025), and hybrid methods (Shi et al., 2020b; Wen et al., 2024; Li et al., 2025b). Voxel-based methods offer good computational efficiency and the ability to represent regular structures, but they tend to lose fine-grained geometric information and have limited performance when detecting small-scale objects. Point-based methods directly extract the features of the point cloud prototypes and retain the original geometric information and spatial structure. Point-voxel methods combine the advantages of the two approaches mentioned above. PVAFN (Li et al., 2025b) introduces keypoint-based geometric details, voxel-wise local regularity, and Bird's-Eye-View (BEV)-level global semantics to enhance feature extraction and fusion. These

methods perform well in general scenarios, but they rely on stable category semantics and appearance patterns and are difficult to be directly applied in industrial defect detection tasks.

3D defect detection in industrial scenarios. Compared to traditional industrial defect detection methods, point cloud-based detection can more accurately capture surface deformations and structural anomalies, thereby improving detection accuracy. First, for feature-enhancing methods (Zhao et al., 2023; Cheng et al., 2026; Li et al., 2025a), the design of multi-scale features and the introduction of attention mechanisms improve the performance to detect small-scale defects. However, these methods still rely on explicit appearance and semantic features, limiting the adaptability to defects with diverse morphologies and no fixed patterns. Second, methods based on geometric modeling (Cao et al., 2025; Tao et al., 2024; Lentsch et al., 2024) utilize geometric information such as normal vectors, curvatures, and local surface fitting to describe the defect area. While these methods can capture structural changes, they are typically used as auxiliary features and lack deep integration with deep learning frameworks (Yu & Zhao, 2026). Furthermore, anomaly detection methods (Liang et al., 2025; Cheng et al., 2025; Cao et al., 2024) treat defect detection as an anomaly detection problem, modeling the normal data distribution to identify abnormal areas without the involvement of anomalous samples. These methods have advantages in unsupervised or semi-supervised scenarios but still fall short in terms of precise localization and object-level detection (such as 3D bounding boxes). Overall, existing methods generally lack systematic analysis of the structural abnormal features of the object, which limits the performance in complex industrial scenarios.

Structure-aware representation learning. The spatial structure information plays a crucial role in 3D detection tasks. Some studies attempt to model the geometric relationships between points by constructing local neighborhood relationships or graph structures, such as using KNN (Jiang et al., 2026; Fu et al., 2025) and GNN (Zhang et al., 2025; Vo et al., 2025) for feature propagation. In the field of representation learning, contrastive learning is widely used to improve feature discriminability. Some researches (Xia et al., 2023; Zhang et al., 2022; 2024; Li et al., 2026) apply this approach to point cloud tasks, optimizing the feature space structure by bringing similar samples closer and different samples farther apart. GPAD (Li et al., 2026) incorporates geometric attributes as auxiliary features into the visual representation to learn the differences in feature distributions between normal and anomalous samples. Nevertheless, the existing methods focus on semantic consistency and instance-level consistency, with few addressing the modeling of local structural patterns. It is worth noting that although these methods utilize structural information to a certain extent, they usually have the following limitations: (1) Structural features are not explicitly modeled as core detection criteria. (2) There is a lack of unified description of structural continuity and structural mutations. (3) They fail to effectively integrate structural information into the detection head phase.

3 METHOD

3.1 OVERVIEW

The defect detection in industrial point clouds has essential differences from traditional 3D object detection tasks. Industrial defects generally do not have stable semantic structures, exhibit diverse appearances, and are extremely small in scale. However, the defects often manifest as abnormal variations in local geometric structures, such as the disruption of surface continuity, local protrusions or depressions, and abrupt changes in normal vectors. Robust detection is difficult to achieve by relying on semantic or appearance features.

Therefore, we propose the geometric metric encoding based on topological consistency for 3D industrial defect detection, which transforms the defect detection problem into the geometric structure anomaly recognition problem. The detection branch of GMetric-3D adopts the general 3D detection supervision paradigm, aiming to predict the location and boundary information of defects in the point cloud. As shown in Fig. 2, the proposed method incorporates geometric metric modeling and topological consistency representation, enabling the network to explicitly learn the local geometric relationships in the point cloud.

Specifically, for the voxel set divided from the input point cloud $\mathcal{P} = \{p_i \in \mathbb{R}^3\}_{i=1}^{N}$, the model first extracts basic 3D features by using the sparse convolutional backbone network. Subsequently, the geometric metric encoding is introduced to construct the local neighborhoods, extract geometric

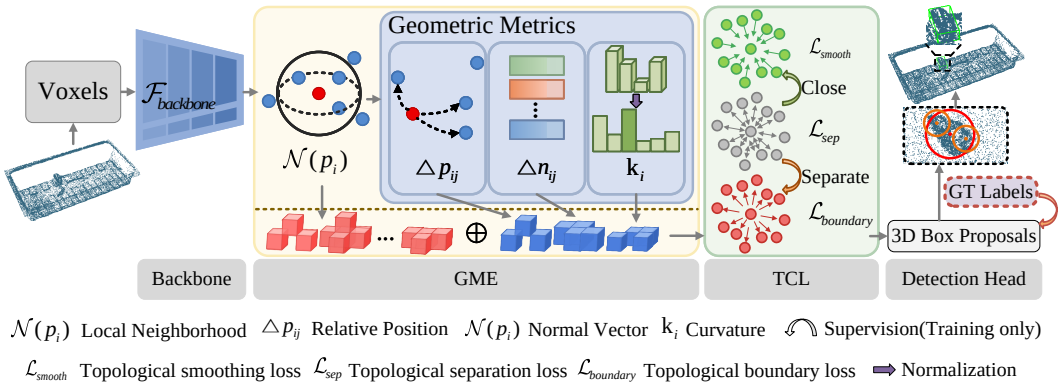


Figure 2: Illustration of the proposed GMetric-3D. At the phase of geometric metric encoding (GME), we first extract the local structural patterns of the point cloud to construct the local neighborhoods. Then, in the geometric metrics extraction, we integrate the geometric metrics containing structural information to capture the local structural features. At the phase of topological consistency learning (TCL), we constrain the topological consistency of the normal regions and enhance the structural differences of the defective regions simultaneously, thereby effectively improving the discriminative ability of the feature space.

metric features, and fuse them with the original features. Based on this, the topological consistency learning is applied to impose constraints on the feature space, ensuring that normal regions maintain structural consistency while defect regions exhibit significant differences. In addition, we introduce an overall loss that enhances the detection performance of the method by constraining the structural relationships in the feature space.

3.2 Geometric Metric Encoding

Sparse 3D backbone. This paper adopts a voxel-based sparse convolutional network from AIFAD (Yu et al., 2026) as the backbone network to efficiently extract large-scale point cloud features. The raw point cloud $\mathcal{P}$ is first divided into the regular voxel grid, and the sparse tensor representation is constructed. Then, spatial features are gradually extracted through multiple layers of sparse 3D convolutions:

$$\mathbf{F} = \mathcal{F}_{backbone}\{\mathcal{P}\} \tag{1}$$

where $\mathbf{F}$ represents the feature representation corresponding to each voxel.

Local neighborhood construction. The traditional feature extraction methods based on sparse convolution primarily aggregate local information through fixed receptive fields. The feature representations extracted by these methods focus on the geometric and semantic distribution of objects but lack the ability to extract explicit structural relations. In particular, when detecting defects that are small-scale or low-contrast defects, these methods fail to capture fine-grained structural variations.

Therefore, in order to extract the local geometric relationships of the point cloud, we introduce the local neighborhood construction. For the given points, we construct the local neighborhood using the fixed spatial offsets $\varphi(\cdot)$ based on the rule-based voxel grids:

$$\mathcal{N}(p_i) = \{p_j | j \in \varphi(p_i, k)\} \tag{2}$$

where k represents the number of grid-based neighborhoods. The construction method focuses on the local region and avoids the problem of global interference.

Geometric metrics extraction. To more effectively describe the structural relationships within a neighborhood, we extract the geometric metrics that integrate multiple pieces of geometric information to capture local structural features from various perspectives. First, we encode the relative position relationship between neighborhood points and the center point:

$$\Delta p_{ij} = p_j - p_i \tag{3}$$

where the relative position relationship represents the spatial distribution pattern of points within the neighborhood. For the raw point cloud in the region bordering the defect object, Δp_{ij} shows a distinct offset.

Second, the normal vector information further captures the direction changes of the surface. We find that the raw point clouds usually do not contain explicit normals. Therefore, we predict the normal vector representations from point features through the feature mapping function $n_i = \psi(\mathbf{F_i})$. The normal vector difference within the neighborhood is obtained by

$$\Delta n_{ij} = n_j - n_i \tag{4}$$

In particular, the normal vector information effectively captures the significant changes on the defect surface.

Finally, we introduce curvature information to further quantify local variations in surface shape. The features are calculated by constructing a covariance matrix:

$$C_i = \frac{1}{k} \sum_{j \in \mathcal{N}(i)} (p_j - \bar{p}_i)(p_j - \bar{p}_i)^T \tag{5}$$

where $\bar{p}_i$ denotes the center of the neighborhood. If the covariance matrices satisfy $\lambda_1 \leq \lambda_2 \leq \lambda_3$, the curvature is defined as $\mathbf{k}_i = \lambda_1/(\lambda_1 + \lambda_2 + \lambda_3)$, where $\mathbf{k}_i$ reflects the degree of curvature of a local surface. The value is close to zero in flat regions, while it significantly increases in raised and depressed areas.

Geometric feature encoding. We encode the metrics such as relative position relationship Δp_{ij}, the normal vector variation Δn_{ij}, and the curvature information $\mathbf{k}_i$ by

$$d_{ij} = \phi([\Delta p_{ij}, \Delta n_{ij}, \mathbf{k}_i]) \tag{6}$$

where $\phi(\cdot)$ is the nonlinear mapping function. Then, the encoded feature d_{ij} is mapped to the feature space, and symmetrical aggregation is performed to obtain the point-level geometric representation:

$$\begin{aligned} \mathbf{s}_i &= \max_{j \in \mathcal{N}_i} (MLP(d_{ij})) \\ \mathbf{F}_i^{'} &= fuse(\mathbf{F}_i, \mathbf{s}_i) \end{aligned} \tag{7}$$

where $fuse(\cdot)$ represents the fusion of point-level geometric representations with original feature representations. The original features provide semantic and contextual information, while the geometric features provide geometric relationships.

3.3 Topological consistency learning

The geometric metric encoding extracts local geometric features, but the network still obtains feature representations that are not discriminative in the absence of effective constraint learning. This is particularly true in industrial settings, where the diverse categories and sparse distribution of defects make models prone to noise and background interference, resulting in unclear boundaries between normal and defective regions in the feature space.

The point cloud of the industrial product surfaces is continuous and structurally stable, exhibiting strong consistency between local regions, while abnormal variations in local geometric relationships are quite pronounced. Therefore, we introduce topological consistency learning to incorporate topological constraints in the feature space. First, for the point-level geometric feature representation $\mathbf{F}_i^{'}$, we define the feature distance between any two points as

$$D_{ij} = \| \mathbf{F}_i^{'} - \mathbf{F}_j^{'} \|_2 \tag{8}$$

where D_{ij} represents the structural similarity between point pairs.

Topological smoothing loss. The point clouds in normal regions exhibit a continuous and smooth structure with stable local topological relationships. In contrast, the topological relationships are often disrupted in defective regions. Therefore, we enhance topological structure consistency by

constraining the similarity of features within local neighborhoods, aiming to constrain feature continuity within regions with the same local topological relationship. Specifically:

$$\mathcal{L}_{smooth} = \sum_{i} \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \| \mathbf{F}_i^{'} - \mathbf{F}_j^{'} \|_2^2 \tag{9}$$

The constraint makes the features in the normal regions smoother and suppresses the interference of local noise on feature learning.

Topological separation loss. Similar to contrastive learning (Wei et al., 2025), we introduce the topological separation loss to minimize the distance between positive sample pairs and incorporate a separation constraint to ensure that abnormal structural regions are separated in the feature space. This loss is defined as

$$\mathcal{L}_{sep} = \sum_{(i,j) \in \mathcal{S}} \max(0, m - \| \mathbf{F}_i^{'} - \mathbf{F}_j^{'} \|_2)^2 \tag{10}$$

where $\mathcal{S}$ represents the set of inconsistent point pairs, and m denotes the separation parameter. The inconsistent point pairs are not directly defined by the manually labeled defects, but are automatically generated based on the local geometric differences.

Topological boundary loss. Due to the discreteness of point clouds and the smoothing operation of features, the detector tends to ignore boundary information, leading to issues such as detection box offset, blurred boundaries, and missed detection of small defects. Therefore, we introduce topological boundary loss to explicitly enhance the perception ability of the detector to perceive structural boundaries. Topological boundary loss aims to constrain the distance relationships in the feature space to be consistent with the actual geometric changes, and avoids the network destroying the actual structural transitions due to excessive smoothing operations. The definition of topological boundary points is

$$\mathcal{B} = \{p_i | \max_{j \in \mathcal{N}(i)} D_{ij} > \tau\} \tag{11}$$

where τ represents the threshold. $\mathcal{B}$ effectively locates regions of structural mutation based on differences in neighborhood features. We constrain the feature differences between boundary points and their neighboring areas:

$$\mathcal{L}_{boundary} = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{N}(i)} w_{ij} \cdot D_{ij} \tag{12}$$

where $w_{ij} = \frac{D_{ij}}{\sum_{j \in \mathcal{N}(i)} D_{ij}}$ denotes weights based on structural differences.

Overall loss. The overall loss $\mathcal{L}$ of GMetric-3D contains the object localization loss $\mathcal{L}_{det}$ supervised by L1 loss, the topology smoothing loss $\mathcal{L}_{smooth}$, the topological separation loss $\mathcal{L}_{sep}$, and the topological boundary loss $\mathcal{L}_{boundary}$. The overall loss $\mathcal{L}$ is

$$\mathcal{L} = \mathcal{L}_{det} + \lambda_1 \mathcal{L}_{smooth} + \lambda_2 \mathcal{L}_{sep} + \lambda_3 \mathcal{L}_{boundary} \tag{13}$$

where λ_1, λ_2, and λ_3 are the coefficients to balance the loss contributions. The topological smoothing loss maintains the topological continuity within normal regions, the topological separation loss enhances the feature distance between normal and defect regions, and the topological boundary loss preserves the geometric mutation at defect boundaries.

4 Experiments

4.1 Experimental details

Implementation details. The GMetric-3D framework we propose is trained in an end-to-end manner and uses the ADAM optimizer. All the detectors were trained on a single RTX 3090 GPU. The network was trained with a learning rate of 3e-3, which was decayed to 1e-4 in a cosine annealing with a decay weight of 0.01. We trained the entire network with a batch size of 2 for 130 epochs. In the implementation of GMetric-3D, we employed data augmentation strategies including global flipping, global scaling, global rotation, GT sampling (Yan et al., 2018), and translation augmentations. We also introduce the parameters, floating-point operations (FLOPs), and inference speed for all compared methods to provide a more detailed analysis of complexity and runtime efficiency.

Table 1: Performance of our GMetric-3D and the state-of-the-art methods on the CPS3D-Det validation set. The results are evaluated by the mean average precision at 11 and 40 recall positions. The best and second-best results are highlighted in red and blue, respectively.

Method	Type	3D AP $IoU = 0.5(R11)$	BEV AP $IoU = 0.5(R11)$	3D AP $IoU = 0.5(R40)$	BEV AP $IoU = 0.5(R40)$	#Params	FLOPs	Infer Speed
Part-A^2 (Shi et al., 2020b)	Point-Voxel	17.11	16.24	16.37	15.43	63.79M	82.78G	199ms
Part-A^2-free (Shi et al., 2020b)	Point-Voxel	13.93	10.91	14.03	9.84	57M	5.85G	229ms
PV-RCNN (Shi et al., 2020a)	Point-Voxel	24.04	24.27	20.40	21.04	13.05M	110.85G	541ms
PointPillars (Lang et al., 2019)	Point-based	30.57	29.36	28.14	26.09	4.81M	9.95G	32ms
CenterPoints (Yin et al., 2021)	Point-based	29.40	33.40	28.71	31.46	7.76M	94.03G	87ms
Fade3D (Ye et al., 2025)	Point-based	19.74	17.72	13.02	10.28	5.30M	76.99G	25ms
PillarNet (Shi et al., 2022)	Voxel-based	35.04	37.25	31.37	34.36	10.98M	180.15G	166ms
SECOND (Yan et al., 2018)	Voxel-based	34.36	33.57	32.90	31.65	5.3M	76.99G	113ms
SECOND-multihead (Yan et al., 2018)	Voxel-based	34.70	34.99	32.14	32.25	5.58M	87.21G	118ms
VoxelNeXt (Chen et al., 2023)	Voxel-based	40.65	37.47	37.77	34.94	7.45M	0.15G	34ms
VoxelNeXt-anchor (Chen et al., 2023)	Voxel-based	34.54	32.61	31.34	29.45	7.37M	0.22G	125ms
PCNet3D++ (Prasanth et al., 2026)	Voxel-based	26.80	29.80	24.61	27.48	4.36M	19.91G	44ms
AIFAD-2D (Yu et al., 2026)	Voxel-based	**40.67**	**39.69**	**39.47**	**38.41**	4.11M	0.24G	36ms
GMetric-3D (Ours)	Voxel-based	**41.86**	**40.36**	**40.90**	**37.20**	7.14M	3.81G	32ms

Table 2: Performance of GMetric-3D and the state-of-the-art methods on the KITTI validation set. The results are evaluated by the mAP at 40 recall positions with 0.5 of IoU.

Method	Type	Car - 3D Detection			Car - BEV Detection			Ped. - 3D Detection			Ped. - BEV Detection			Cyc. - 3D Detection			Cyc. - BEV Detection		
		Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard
Part-A^2 (Shi et al., 2020b)	Point-Voxel	97.80	93.72	93.20	97.83	93.81	93.54	61.58	53.80	48.25	64.86	57.46	52.43	87.45	73.59	69.17	91.67	78.43	73.98
Part-A^2-free (Shi et al., 2020b)	Point-Voxel	95.86	93.75	91.77	95.88	93.89	93.55	68.30	62.28	55.92	72.15	66.42	60.83	84.63	73.14	68.64	88.50	77.24	72.49
PV-RCNN (Shi et al., 2020a)	Point-Voxel	97.90	94.43	94.17	97.92	94.49	94.27	58.92	51.91	47.39	62.28	56.00	51.91	86.41	71.13	66.96	88.86	75.55	71.05
Fade3D (Ye et al., 2025)	Point-based	95.50	94.00	92.93	95.52	94.14	93.53	43.88	39.59	36.23	53.53	48.74	45.55	79.04	58.31	54.54	87.48	65.55	61.79
PointPillars (Lang et al., 2019)	Point-based	94.57	93.08	92.13	94.62	93.36	92.97	37.67	34.76	31.86	46.07	43.23	39.94	74.50	58.49	54.64	78.48	63.82	59.58
Point-RCNN (Shi et al., 2019)	Point-based	96.23	92.90	90.58	96.26	94.82	90.65	64.01	55.09	48.29	67.67	59.53	51.93	89.55	71.00	67.68	93.68	74.49	71.04
VoxelNeXt (Chen et al., 2023)	Voxel-based	95.50	94.00	91.52	95.52	94.25	93.48	56.85	52.75	47.75	60.81	57.06	52.52	78.42	59.07	55.35	81.27	62.24	58.48
CT3D (Sheng et al., 2021)	Voxel-based	95.01	89.96	87.89	95.00	90.36	89.99	49.00	42.55	37.93	53.94	47.21	43.17	72.04	52.44	49.67	72.85	54.71	51.21
PillarNet (Shi et al., 2022)	Voxel-based	94.73	92.99	90.88	94.81	93.43	92.93	41.99	37.58	34.58	51.86	47.25	44.13	78.52	59.03	54.84	84.29	64.52	59.82
SECOND (Yan et al., 2018)	Voxel-based	94.99	93.64	92.97	95.03	93.83	93.45	45.66	42.18	37.64	52.66	49.25	44.59	80.51	66.12	62.18	83.14	68.95	64.97
PCNet3D++ (Prasanth et al., 2026)	Voxel-based	94.16	87.01	84.31	97.67	91.76	89.20	46.07	39.82	35.84	53.30	47.59	44.16	70.35	52.19	48.94	75.94	57.16	53.68
AIFAD (Yu et al., 2026)	Voxel-based	95.06	92.77	90.48	95.18	93.15	92.22	55.46	50.30	45.10	61.13	56.16	51.59	80.56	61.29	57.62	83.13	66.06	62.34
GMetric-3D(Ours)	Voxel-based	97.04	91.01	90.32	97.18	93.15	90.80	56.50	50.57	45.89	62.63	57.26	52.77	82.96	64.30	60.35	85.49	68.48	64.39

Datasets. The high-resolution point cloud defect detection plays a crucial role in the fields of precision manufacturing and processing, where it is often necessary to identify defects in the complex structures of industrial products. The CPS3D-Det dataset (Yu et al., 2026), with a total number of 1640 point cloud data, consists of 928 training samples, 312 testing samples, 170 normal samples, and 230 more challenging abnormal samples. The samples have very high point cloud quality, with the spatial resolution of 0.0025 mm. The range of the point cloud on the X-axis is [0, 776 mm], on the Y-axis it is [-368 mm, 368 mm], and on the Z axis it is [-120 mm, 40 mm]. All the samples are from real industrial scenarios, reflecting the environmental changes in industrial production.

Similar to the experimental schemes in 3D defect detection methods (Yu et al., 2026; Yu & Zhao, 2026), we choose the KITTI (Geiger et al., 2012) dataset to validate the effectiveness of GMetric-3D. The KITTI dataset is one of the most popular datasets in the field of 3D detection. There are 7481 training samples and 7518 testing samples, where the training samples are generally divided into the train split (3712 samples) and the val split (3769 samples). We evaluate our model on this large-scale dataset to further validate the effectiveness of our proposed method.

4.2 Main Results

Evaluation on CPS3D-Det dataset. We compare our GMetric-3D with existing state-of-the-art methods on the CPS3D-Det dataset and present the quantitative results in Tab. 1. For the point-voxel methods, Part-A^2 (Shi et al., 2020b) and PV-RCNN (Shi et al., 2020a) exhibit poor performance on evaluation metrics. This is because high-precision defects have complex appearances and a small number of points, and the rich contextual information is lost by the ROI global pooling operation, which leads to erroneous object centers within each 3D region proposal. The point-based methods directly sample the raw point cloud to extract spatial features of the sparser defects. In contrast, our GMetric-3D models the local neighborhood to extract geometric structure features and fuses them with the original features, preserving the original spatial structure. Therefore, GMetric-3D outperforms PointPillars (Lang et al., 2019), CenterPoints (Yin et al., 2021), and Fade3D (Ye et al., 2025) in terms of performance, achieving 11.29%, 12.46%, and 22.12% improvements in 3D average accuracy at 11 recall thresholds, respectively. Among voxel-based methods, GMetric-3D achieves the best

Table 3: Ablation study of GMetric-3D. We conduct the ablation studies on the effects of relative position, normal vector, and curvature in GME and the topology consistency loss, structure separation loss, and structure boundary loss in TCL.

GME			TCL			3D AP $R40$	BEV AP $R40$
Relative position	Normal vector	Curvature	$\mathcal{L}_{smooth}$	$\mathcal{L}_{sep}$	$\mathcal{L}_{boundary}$		
✗	✗	✗	✗	✗	✗	22.74	24.97
✗	✗	✗	✓	✓	✓	24.31	25.30
✓	✓	✓	✗	✗	✗	29.39	29.09
✓	✗	✗	✓	✗	✗	24.18	27.06
✗	✓	✗	✗	✓	✗	23.49	26.40
✗	✗	✓	✗	✗	✓	24.56	27.42

Table 4: Effects of different backbone schemes. The evaluation metrics follow the 40 recall positions.

Backbone	3D AP IoU=0.5	BEV AP IoU=0.5	3D AP IoU=0.7	BEV AP IoU=0.7	Latency(ms)
SECOND (Yan et al., 2018)	23.75	21.04	9.48	3.45	39
VoxelNeXt (Chen et al., 2023)	31.80	27.93	15.37	5.54	43
GMetric-3D(Ours)	40.90	37.20	28.68	12.59	32

performance and reduces the number of parameters. Compared to the second-best method, AIFAD (Yu et al., 2026), GMetric-3D improves 3D AP and BEV AP by 1.19% and 0.67% respectively, at 11 recall threshold. Our GMetric-3D focuses on the topological consistency and the spatial geometric metric, preserving the original geometric information and spatial structure.

Evaluation on KITTI dataset. In order to further verify the performance of our benchmark on other bounding-box-based datasets, we conduct experiments on the KITTI validation set. As shown in Tab. 2, compared with the state-of-the-art methods, our GMetric-3D has competitive performance. Our method outperforms other methods on several metrics, this is because the objects of the KITTI dataset usually have clear semantic categories and stable appearance structures, and the background regions have very intense appearance curves to reduce the performance of the methods. For the 3D and bird's-eye-view (BEV) detections of the car and pedestrian classes, our method also achieves competitive performance on the easy, moderate, and difficult levels in the comparison of voxel-based methods. This validates the motivation of our research.

4.3 Ablation Study

Module ablation. To more effectively verify the effectiveness of each component in our proposed GMetric-3D, we present performance comparison results for six different configurations in Tab. 3. As shown in the first three rows, it can be seen that we introduce geometric metric encoding (GME) and topological consistency learning (TCL) into the baseline, and the performances are all improved. GME constructs the local neighborhood relationships between point pairs and enhances the discriminative local structural features by combining the metrics of relative positions, normal vectors, and curvature. Furthermore, as shown in the last three rows, we set up three different strategies to verify the relationship between the components in GME and TCL. The results indicate that introducing each component individually yields significant performance improvements over the baseline. This is because the geometric metrics enhance the feature space, enabling the detector to focus on the structural mutation features of the defects, thereby improving the performance of the network in locating and recognizing defects.

Effect of backbone schemes. We explore the impact of the backbone network on performance and conduct ablation studies using the backbone networks from SECOND (Yan et al., 2018) and VoxelNeXt (Chen et al., 2023). As shown in Tab. 4, compared to SECOND and VoxelNeXt, the proposed GMetric-3D achieves improvements of 13.31%, 19.2% and 9.1%, 17.15% in AP_{3D} at IoU of 0.5 and 0.7, respectively. This is because SECOND and VoxelNeXt focus on semantic features

Table 5: Effects of the loss balance coefficient. The evaluation metrics follow the 40 recall positions.

Coefficient	3D AP $R40$	Recall (%) IoU=0.5
λ_1=0.1, λ_2=0.1, λ_3=0.2	24.31	77.47
λ_1=0.2, λ_2=0.2, λ_3=0.4	25.13	78.92
λ_1=0.4, λ_2=0.4, λ_3=0.4	22.59	79.93
λ_1=0.4, λ_2=0.4, λ_3=0.8	29.79	79.76
λ_1=0.8, λ_2=0.8, λ_3=1.0	27.99	78.92
λ_1=1.0, λ_2=1.0, λ_3=1.0	31.82	82.46

Table 6: Effects of the number k of grid-based neighborhoods. The evaluation metrics follow the 40 recall positions.

Neighborhood number	3D AP $R40$	Recall (%) IoU=0.5
k = 6	40.90	80.78
k = 18	29.22	76.56
k = 26	25.94	73.52

Table 7: Effects of the different fusion methods. The evaluation metrics follow the 11 recall positions.

Fusion method	3D AP $IoU_{0.5}$	BEV AP $IoU_{0.5}$
Backbone		
+ Transformer	*31.89*	*29.77*
+ GME	*41.86*	*40.36*

Table 8: Effects of the different detection head. The evaluation metrics follow the 40 recall positions.

Detection head	3D AP $IoU_{0.5}$	BEV AP $IoU_{0.5}$
TCL		
+ Anchor-based	*15.70*	*18.61*
+ Anchor-free	*40.90*	*37.20*

and appearance patterns, lacking the analysis of the essential representation of defects. In contrast, our GMetric-3D focuses on the abnormal deviations of the local structural patterns, perceiving and utilizing local geometric relationships to significantly improve detection performance.

Ablation of model details. We conduct more ablation experiments to analyze the effects of model details. First, for the loss balance coefficient, as shown in Tab. 5, both detection accuracy and recall increase as the weight coefficient increases. This indicates that by introducing topological smoothing loss, topological separation loss, and topological boundary loss to constrain the similarity of features within the local neighborhood, the proposed method enhances the topological consistency and improves the performance of the network. Then, we present Tab. 6 to explore the influence of the neighborhood number k in Eq. 2 on the performance of the network. Among $k = 6$ (surface neighborhood), $k = 18$ (line and surface neighborhoods), and $k = 26$ (point, line, and surface neighborhoods), the surface neighborhood achieves the best performance. This is because the local geometric information relies on the direct surface neighborhood, while the information carried by the point and line has weaker correlations with the central voxel. Moreover, as shown in Tab. 7 and Tab. 8, there are potential choices of transformer fusion mechanism (Zhao et al., 2021) and anchor-based method (Yan et al., 2018). The transformer fusion mechanism focuses on semantic features and appearance patterns, which presents significant limitations in industrial defect detection scenarios. The anchor-based strategy requires manually defined anchor box sizes, while industrial defects usually do not have stable category semantics. The results demonstrate that the GMetric-3D effectively captures subtle geometric variations, thereby improving the detection effect of the network.

5 Conclusion

Industrial defects are characterized by irregular shapes and subtle geometric variations and manifest as local structural anomalies. The existing 3D detection methods rely on semantic and appearance features, which fail to effectively capture the subtle geometric changes. We propose the geometric metric encoding based on topological consistency (GMetric-3D), focusing on the geometric mutation features and topological consistency of the defects. We integrate the intrinsic attributes of defect occurrence in real industrial scenarios, and topological consistency learning is constructed to preserve topological consistency of defective samples through multiple complementary constraints. The results show that explicitly modeling geometric metrics plays a crucial role in improving the accuracy and robustness in industrial scenarios. This validates the motivation of our research.

AI USE STATEMENT

In this work, we used generative AI tools for grammar checking and spell-checking. We have not used generative AI tools for generating illustrations or writing the main text. We take responsibility for the final content of this work produced with the aid of generative AI.

ETHICS STATEMENT

Our research conducted in the paper conform with the ICLR code of ethics.

REFERENCES

Mujtaba Asad, Waqar Azeem, He Jiang, Hafiz Tayyab Mustafa, Jie Yang, and Wei Liu. 2m3df: Advancing 3d industrial defect detection with multi-perspective multimodal fusion network. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(7):6803–6815, 2025.

Xuanming Cao, Chengyu Tao, and Juan Du. 3d-adcs: Untrained 3d anomaly detection for complex manufacturing surfaces. *Journal of Computing and Information Science in Engineering*, 25:111002, 2025.

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

Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21674–21683, 2023.

Yuqi Cheng, Yunkang Cao, Dongfang Wang, Weiming Shen, and Wenlong Li. Boosting global-local feature matching via anomaly synthesis for multi-class point cloud anomaly detection. *IEEE Transactions on Automation Science and Engineering*, 22:12560–12571, 2025.

Yuqi Cheng, Yihan Sun, Hui Zhang, Weiming Shen, and Yunkang Cao. Towards high-resolution 3d anomaly detection: A scalable dataset and real-time framework for subtle industrial defects. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 3327–3334, 2026.

Yufei Fu, Yuhao Guo, and Hui Hu. Prnet: 3d object detection network-based on point-region fusion. *Applied Sciences*, 15(7):3759, 2025.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 3354–3361. IEEE, 2012.

Chen Jiang, Shuxia Lu, Xianghu Zhou, Tingting Ma, and Junhai Zhai. Knn improved transformer for 3d object detection. *Signal Processing: Image Communication*, pp. 117488, 2026.

Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12697–12705, 2019.

Ted Lentsch, Holger Caesar, and Dariu M Gavrila. Union: Unsupervised 3d object detection using object appearance-based pseudo-classes. *Advances in Neural Information Processing Systems*, 37: 22028–22046, 2024.

Min Li, Jinghui He, Gang Li, Jiachen Li, Jin Wan, and Delong Han. Multimodal industrial anomaly detection via geometric prior. *IEEE Transactions on Circuits and Systems for Video Technology*, 36:2854–2866, 2026.

Yaochen Li, Qiao Li, Cong Gao, Shengjing Gao, Hao Wu, and Rui Liu. Pfenet: Towards precise feature extraction from sparse point cloud for 3d object detection. *Neural Networks*, 185:107144, 2025a.

Yidi Li, Jiahao Wen, Rui Gong, Bin Ren, Wenhao Li, Chen Cheng, Hong Liu, and Nicu Sebe. Pvafn: Point-voxel attention fusion network with multi-pooling enhancing for 3d object detection. *Expert Systems with Applications*, 281:127608, 2025b.

Hanzhe Liang, Guoyang Xie, Chengbin Hou, Bingshu Wang, Can Gao, and Jinbao Wang. Look inside for more: Internal spatial modality perception for 3d anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5146–5154, 2025.

Chuandong Liu, Chenqiang Gao, Fangcen Liu, Pengcheng Li, Deyu Meng, and Xinbo Gao. Hierarchical supervision and shuffle data augmentation for 3d semi-supervised object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23819–23828, 2023.

Kaiqi Liu, Yuanyuan Deng, Jiaxun Tong, and Wei Li. Univoxel: A novel framework for 3-d object detection in autonomous vehicles with multimodal voxel representation. *IEEE Sensors Journal*, 25 (17):33142–33152, 2025.

Thurimerla Prasanth, Ram Prasad Padhy, and B. Sivaselvan. Pcnet3d++: A pillar-based cascaded 3d object detection model with an enhanced 2d backbone. *Image and Vision Computing*, 166:105854, 2026.

Hualian Sheng, Sijia Cai, Yuan Liu, Bing Deng, Jianqiang Huang, Xian-Sheng Hua, and Min-Jian Zhao. Improving 3d object detection with channel-wise transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 2743–2752, 2021.

Guangsheng Shi, Ruifeng Li, and Chao Ma. Pillarnet: Real-time and high-performance pillar-based 3d object detection. In *European Conference on Computer Vision*, pp. 35–52. Springer, 2022.

Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 770–779, 2019.

Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10529–10538, 2020a.

Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2647–2664, 2020b.

Chengyu Tao, Hao Xu, and Juan Du. F2pad: A general optimization framework for feature-level to pixel-level anomaly detection. *arXiv preprint arXiv:2407.06519*, 2024.

Ching-Yu Tseng, Yi-Rong Chen, Hsin-Ying Lee, Tsung-Han Wu, Wen-Chin Chen, and Winston H Hsu. Crossdtr: Cross-view and depth-guided transformers for 3d object detection. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4850–4857. IEEE, 2023.

Thai Anh Vo, Lan Anh Nguyen, Trung Son Doan, and Son Hong Ngo. Gnns-based 3d object detection in autonomous driving: System and empirical evaluation. In *2025 16th International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1874–1879. IEEE, 2025.

Shuhua Wang, Ke Lu, Yang Zhao, Hengsheng Lun, Zehai Niu, and Jian Xue. Vs3d: A vote-based semi-supervised 3d object detection framework for point clouds. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2024a.

Yuqi Wang, Yuntao Chen, Xingyu Liao, Lue Fan, and Zhaoxiang Zhang. Panoocc: Unified occupancy representation for camera-based 3d panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17158–17168, 2024b.

Shun Wei, Jielin Jiang, and Xiaolong Xu. Uninet: A contrastive learning-guided unified framework with feature selection for anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 9994–10003, 2025.

Keli Wen, Nan Zhang, Ge Li, and Wei Gao. Mpvnn: multi-resolution point-voxel non-parametric network for 3d point cloud processing. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2024.

Hai Wu, Chenglu Wen, Shaoshuai Shi, Xin Li, and Cheng Wang. Virtual sparse convolution for multimodal 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 21653–21662, 2023.

Qiming Xia, Jinhao Deng, Chenglu Wen, Hai Wu, Shaoshuai Shi, Xin Li, and Cheng Wang. Coin: Contrastive instance feature mining for outdoor 3d object detection with very limited annotations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6254–6263, 2023.

Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18 (10):3337, 2018.

Zhao Yang, Yinan Shi, Weixiang Xu, Jiangtong Zhu, Haonan Zhang, and Longjun Liu. Object-guided semi-supervised bird's-eye view 3d object detection with 3d box refinement. *IEEE Transactions on Intelligent Transportation Systems*, 27(1):414–432, 2025.

Wei Ye, Qiming Xia, Hai Wu, Zhen Dong, Ruofei Zhong, Cheng Wang, and Chenglu Wen. Fade3d: Fast and deployable 3d object detection for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 26(9):12934–12946, 2025.

Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11784–11793, 2021.

Ruiyun Yu and Ziming Zhao. Dacc: Domain adaptation based on coupling consistency for semi-supervised 3d defect detection. *IEEE Transactions on Multimedia*, 2026.

Ruiyun Yu, Ziming Zhao, and Shi Zhen. Auxiliary information flow for 3d industrial defect detection on ic ceramic package substrate surfaces: Dataset and benchmark. *IEEE Transactions on Circuits and Systems for Video Technology*, 36:7444–7456, 2026.

Taolin Zhang, Sunan He, Tao Dai, Zhi Wang, Bin Chen, and Shu-Tao Xia. Vision-language pre-training with object contrastive learning for 3d scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 7296–7304, 2024.

Xiyu Zhang, Jiayi Ma, Jianwei Guo, Wei Hu, Zhaoshuai Qi, Fei Hui, Jiaqi Yang, and Yanning Zhang. Hypergct: A dynamic hyper-gnn-learned geometric constraint for 3d registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24750–24759, 2025.

Yanan Zhang, Jiaxin Chen, and Di Huang. Cat-det: Contrastively augmented transformer for multi-modal 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 908–917, 2022.

Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16259–16268, 2021.

Xinyue Zhao, Quanzhi Li, Menghan Xiao, and Zaixing He. Defect detection of 3d printing surface based on geometric local domain features. *The International Journal of Advanced Manufacturing Technology*, 125(1):183–194, 2023.

Hanqi Zhu, Jiajun Deng, Yu Zhang, Jianmin Ji, Qiuyu Mao, Houqiang Li, and Yanyong Zhang. Vpfnet: Improving 3d object detection with virtual point based lidar and stereo data fusion. *IEEE Transactions on Multimedia*, 25:5291–5304, 2022.

A APPENDIX

In the appendix, we provide more visual analyses, including the evolution of the training process and the analysis of the feature channels, which further validate the effectiveness of GMetric-3D.

A.1 BALANCE BETWEEN SPEED AND PRECISION

We conduct the balance analysis between speed and precision for GMetric-3D and existing state-of-the-art methods. As shown in Fig. 3, the proposed GMetric-3D achieves the highest accuracy without sacrificing too much efficiency. In terms of speed, it ranks second only to Fade3D (Ye et al., 2025) and is the same as PointPillars (Lang et al., 2019).

A.2 DYNAMIC EVOLUTIONARY DURING TRAINING

As shown in Fig. 4, we visualize the dynamic evolution process of VoxelNeXt (Chen et al., 2023), SECOND (Yan et al., 2018), and GMetric-3D during the training process. VoxelNeXt and SECOND exhibit significant fluctuations during training, while our GMetric-3D remains in a state of steady growth. This indicates that GMetric-3D identifies defect regions by modeling the normal data distribution, which validates our research motivation.

Furthermore, we employ the fixed spatial offsets based on the rule-based voxel grids to construct the local neighborhood. There are three potential strategies to choose from: (1) surface neighborhood ($k = 6$), (2) line and surface neighborhoods ($k = 18$), and (3) point, line, and surface neighborhoods ($k = 26$). As shown in Fig. 5, GMetric-3D achieves optimal performance when $k = 6$. This is because the local geometric information relies on the direct surface neighborhood, while the information carried by the point and line has weaker correlations with the central voxel.

A.3 FEATURE VISUALIZATION

As shown in Fig. 6, we provide 2D and 3D feature visualizations for VoxelNeXt, SECOND, and GMetric-3D. GMetric-3D captures the clear local features of minute defects, while VoxelNeXt and SECOND have limited ability to identify textures in defect regions.

Furthermore, to further analyze the impact of 3D sparse features on detection results, we visualize the mean, standard deviation, mean combined with standard deviation, activation ratio, and importance score for each channel in Fig. 7. By comparing these statistics, it can be seen that channels with higher importance scores tend to exhibit higher activation rates and mean values, while channels with larger standard deviations contribute more inconsistently to the detection results. This indicates that the activation intensity and frequency of features play a dominant role in the detection task, and integrating mean and standard deviation can more comprehensively evaluate the effectiveness of the channels.

A.4 FURTHER WORK

The GMetric-3D has the potential to bring significant benefits to industrial inspection and intelligent manufacturing. By modeling geometric structures in point clouds, the method improves the accuracy and robustness of defect detection, particularly for small-scale and subtle defects that are difficult to identify using conventional approaches. Furthermore, in edge cases involving extremely subtle defects or noisy point clouds, sufficient human oversight mechanisms become particularly important. In summary, the proposed method provides a promising direction for improving industrial inspection through structure-aware learning, while careful consideration of deployment strategies, data diversity, and human oversight is necessary to ensure its safe and responsible use.

Based on our work, there is still a broad scope for improvement and exploration. First, the proposed method relies on local neighborhood construction, which may be sensitive to variations in point density. Second, the structural modeling is based on geometric cues and has not yet incorporated higher-level contextual information.

In future works, we will explore more efficient neighborhood construction strategies and investigate adaptive structural modeling mechanisms. Furthermore, extending the proposed framework to semi-

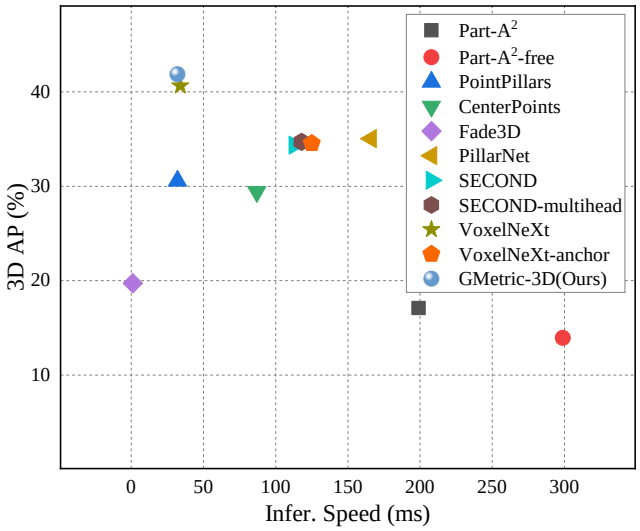
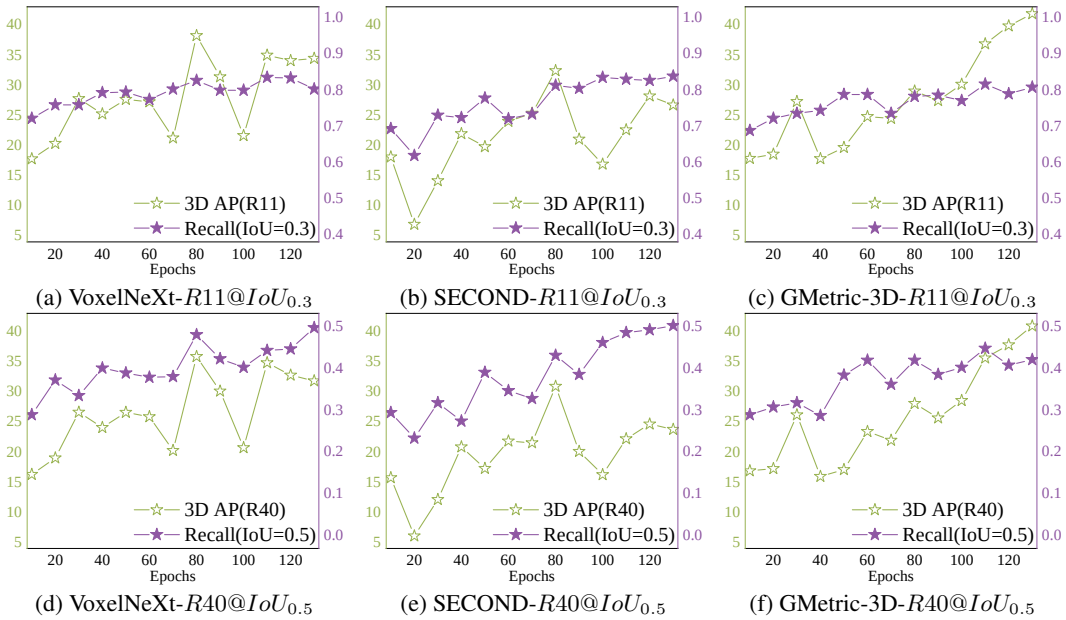


Figure 3: The scatter plot results on the CPS3D-Det dataset. The horizontal axis represents the model inference speed, and the vertical axis represents 3D average precision. Our GMetric-3D demonstrated the best performance.

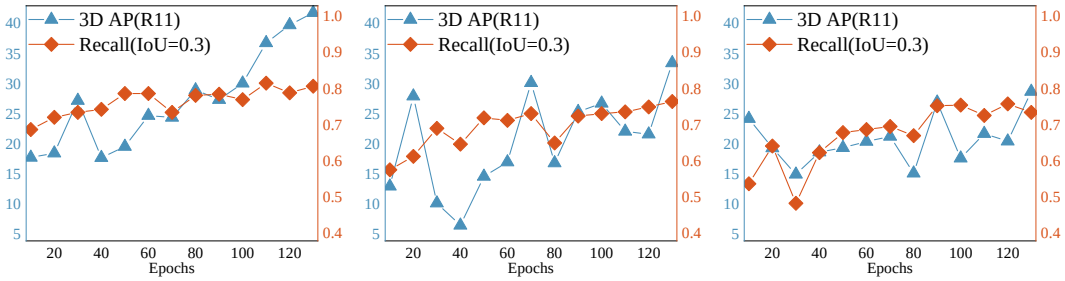


(a) VoxelNeXt-$R11@IoU_{0.3}$ (b) SECOND-$R11@IoU_{0.3}$ (c) GMetric-3D-$R11@IoU_{0.3}$

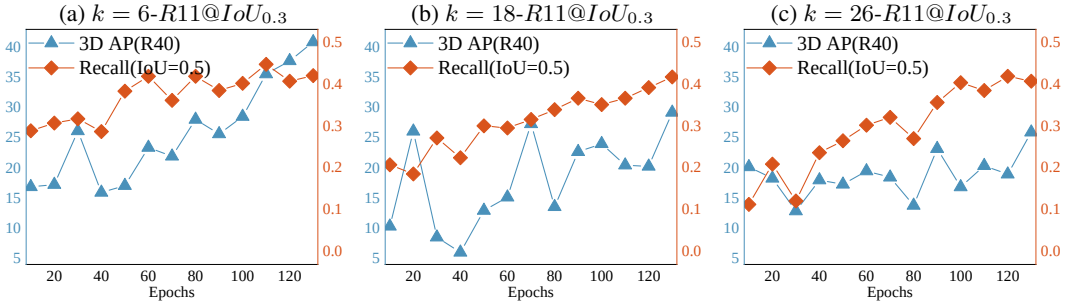
(d) VoxelNeXt-$R40@IoU_{0.5}$ (e) SECOND-$R40@IoU_{0.5}$ (f) GMetric-3D-$R40@IoU_{0.5}$

Figure 4: The effects of different backbone networks among VoxelNeXt, SECOND, and GMetric-3D.

supervised and unsupervised settings for unknown defect detection is a promising direction. We also plan to integrate global context and multi-scale structural representations to further enhance the robustness and generalization ability of the model.

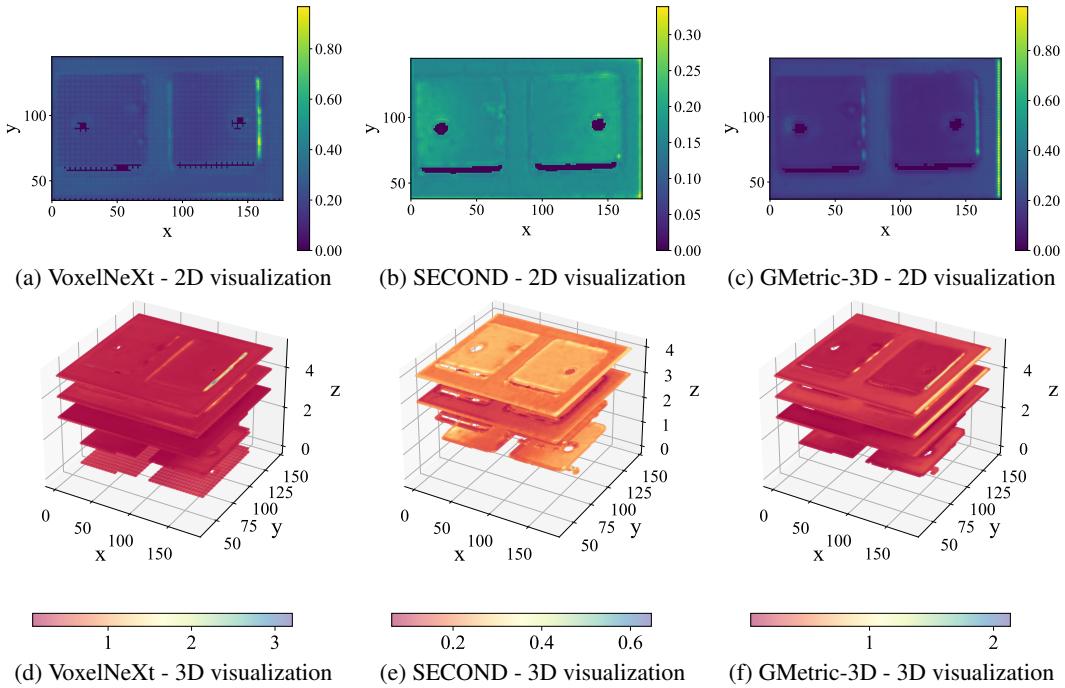


(a) $k = 6$-$R11@IoU_{0.3}$ (b) $k = 18$-$R11@IoU_{0.3}$ (c) $k = 26$-$R11@IoU_{0.3}$



(d) $k = 6$-$R40@IoU_{0.5}$ (e) $k = 18$-$R40@IoU_{0.5}$ (f) $k = 26$-$R40@IoU_{0.5}$

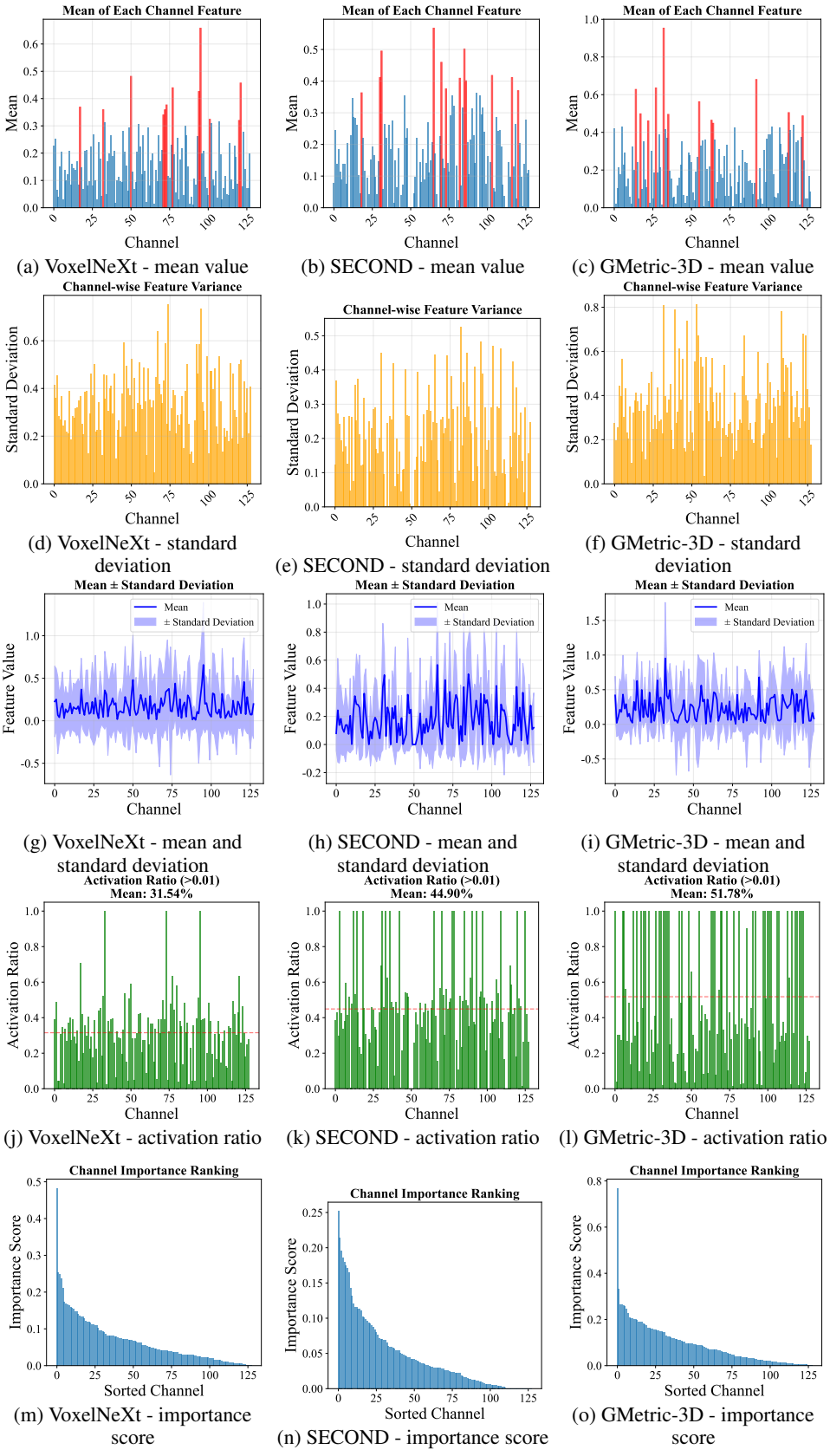
Figure 5: The effects of different neighborhood number k among 6, 18, and 26.



(a) VoxelNeXt - 2D visualization (b) SECOND - 2D visualization (c) GMetric-3D - 2D visualization

(d) VoxelNeXt - 3D visualization (e) SECOND - 3D visualization (f) GMetric-3D - 3D visualization

Figure 6: The feature visualization of 2D and 3D features among VoxelNeXt, SECOND, and GMetric-3D.



(a) VoxelNeXt - mean value (b) SECOND - mean value (c) GMetric-3D - mean value

(d) VoxelNeXt - standard deviation (e) SECOND - standard deviation (f) GMetric-3D - standard deviation

(g) VoxelNeXt - mean and standard deviation (h) SECOND - mean and standard deviation (i) GMetric-3D - mean and standard deviation

(j) VoxelNeXt - activation ratio (k) SECOND - activation ratio (l) GMetric-3D - activation ratio

(m) VoxelNeXt - importance score (n) SECOND - importance score (o) GMetric-3D - importance score

Figure 7: The channel analysis among VoxelNeXt, SECOND, and GMetric-3D.