

Received 15 June 2025, accepted 25 June 2025, date of publication 3 July 2025, date of current version 10 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3585843

RESEARCH ARTICLE

Survey on 3D Anomaly Detection Regarding Defect Size: Tradeoff Between Accuracy and Efficiency With Future Research Direction

A-SEONG MOON^{ID}, KUNHO LEE^{ID}, AND JAESUNG LEE^{ID}

Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, Republic of Korea

Corresponding author: Jaesung Lee (curseor@cau.ac.kr)

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korea government [Ministry of Science and ICT (MSIT)] [RS-2021-II211341, Artificial Intelligence Graduate School Program, Chung-Ang University]; and in part by the Chung-Ang University Graduate Research Scholarship (Academic Scholarship for the College of Biotechnology and Natural Resources) in 2023.

ABSTRACT Recent advancements in neural networks has enhanced image anomaly detection in different industries such as manufacturing, reduced human intervention, and boosted productivity and quality. However, there are still some challenges in using point cloud and RGB data for anomaly detection across varying defect sizes. This study explores such challenges and highlights the importance of high-quality data for effective training. It also highlights the trade-off between detection accuracy and inference speed, where faster models tend to sacrifice performance on smaller defects due to the use of general loss functions that do not sufficiently capture fine-grained anomalies. This work suggests that future research should focus on improving the models that efficiently detect small defects while ensuring high-speed inference and bridging the gap between accuracy and efficiency for real-world industrial applications.

INDEX TERMS 3D anomaly detection, defect size analysis, industrial inspection, multimodal learning, efficiency-accuracy trade-off.

I. INTRODUCTION

With advancements in artificial intelligence (AI), quality assurance in manufacturing has become increasingly automated, thereby reducing the need for human intervention. Automation, particularly in environments that are hostile or hazardous to humans, not only enhances productivity but also contributes to worker welfare by minimizing direct exposure to dangerous conditions.

Initially, AI technologies focused on identifying surface-level defects using two-dimensional (2D) image anomaly detection. These methods significantly improve the speed and reliability of quality assurance in production lines. However, as products become more complex and require inspections beyond surface-level anomalies, these techniques must be

extended to three-dimensional (3D) data. This transition from 2D to 3D anomaly detection was crucial, as it enabled the identification of defects in intricate geometries, internal structures, and surface irregularities that 2D methods could not detect [1].

In recent years, neural networks have significantly advanced in learning from complex data formats, including high-dimensional, temporal, and spatial data. Anomaly detection has achieved remarkable success, prompting researchers to explore both 2D and 3D data. However, the researchers often face challenges related to the high computational demands of processing large volumes of data and the difficulty of obtaining high-quality, labeled datasets in industrial settings. These challenges have led to the exploration of new approaches for efficient and robust 3D anomaly detection.

Although research on defect detection has been actively evolving, one crucial aspect that has not been thoroughly

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen^{ID}.

investigated is the impact of defect size on detection performance. Most existing studies focus on improving detection accuracy through advanced architectures and feature extraction techniques but do not explicitly analyze how variations in defect size affect model performance. Understanding this relationship is essential for developing more robust and scalable anomaly detection systems, particularly for applications where defects can vary significantly in size.

Recent work by Rolih et al. [2] demonstrated that defect size significantly affects anomaly detection models. Larger defects are more readily identifiable, whereas smaller defects are often overlooked because of their subtle appearance and the limitations of feature extraction processes. This study aims to bridge this gap by systematically analyzing how defect size influences detection performance in anomaly detection. By evaluating detection accuracy across various defect scales and comparing different approaches, we seek to provide insights that can contribute to the development of more effective anomaly detection frameworks for industrial applications.

In this study, we systematically analyze the impact of defect size on multimodal anomaly detection using the MVTec 3D-AD dataset and make the following key contributions:

- **Comparative Study of Existing Approaches:** We evaluated various anomaly detection methods, including reconstruction, representation, and student-teacher network-based approaches, to assess their effectiveness across different defect sizes.
- **Systematic Analysis of Defect Size Influence:** We conducted extensive experiments on the MVTec 3D-AD dataset to investigate how defect size affects the performance of anomaly detection models, particularly for 3D point-cloud and RGB data.
- **Identification of Key Challenges and Future Directions:** We highlight the limitations of current methods in handling small-scale defects and propose future research directions to enhance the robustness and efficiency of multimodal defect detection models.

Through this work, we aim to provide valuable insights into optimizing defect detection models by considering defect size as a critical factor, ultimately contributing to the advancement of real-world industrial applications.

II. RELATED WORK

Defect detection is inherently linked to anomaly detection, as both tasks involve identifying instances that deviate from expected normal patterns. In industrial settings, defects often manifest as anomalies—uncommon deviations from standard manufacturing conditions. As a result, many anomaly detection techniques have been adapted for defect detection, leveraging the assumption that defects are rare and significantly different from normal samples. Given the challenges of obtaining labeled defect samples,

unsupervised anomaly detection methods have gained traction, providing effective solutions by learning the distribution of normal samples and identifying deviations as potential defects.

Anomaly detection has evolved into a fundamental pillar of industrial automation, where detecting defects with precision and real-time efficiency is crucial for minimizing downtime and ensuring product quality. Over the years, this field has progressed from basic image-based approaches to advanced frameworks that integrate multimodal data sources, including RGB images, depth maps, and sensor signals. As outlined in Table 1, recent studies have explored various methods, including representation learning, reconstruction-based approaches, and knowledge distillation, to address these challenges in various type of defect detection. These advancements have been particularly catalyzed by the challenges posed by industrial variability, where defect types range from subtle surface scratches to structural anomalies invisible in single-modal inputs.

The year 2022 marks a pivotal turning point in anomaly detection research. During this period, researchers began addressing not only visual anomalies but also those embedded in sensor signals and heterogeneous data environments. This shift was driven by the increasing complexity of industrial systems, where anomalies can manifest in various forms, including structural defects, operational faults, and system failures. As a result, recent studies have extended traditional image-based anomaly detection to encompass multimodal approaches that integrate visual, thermal, acoustic, and vibration data, enabling more comprehensive fault diagnosis.

One such study, Jiang et al. [3], tackled the challenge of industrial multi-sensor signal anomaly detection, emphasizing the difficulties posed by distinguishing regular, defect, and abnormal signals in complex environments. Their study highlighted how abnormal signals, caused by factors like temperature fluctuations and sensor noise, can obscure defect detection and lead to false positives. To mitigate this, they explored normal signal reconstruction, abnormal signal modeling, and statistical approaches. Their proposed LSTM-autoencoder framework leveraged spatiotemporal features to distinguish anomalies more effectively, improving detection accuracy and reducing false positives in industrial applications.

Expanding on the need for robust anomaly detection in heterogeneous industrial datasets, Jeong et al. [18] proposed a two-stage deep anomaly detection (T-DAD) framework designed to handle diverse time-series signals. Their approach addressed the challenge of integrating operation cycle signals and sensor signals, which exhibit distinct temporal patterns. By leveraging a sequential anomaly selection and refinement process, their method effectively utilized signal-specific characteristics to enhance detection accuracy. Their work contributed to the development of data-driven anomaly detection strategies that are robust to imbalanced labels and unpredictable anomaly patterns in industrial settings.

TABLE 1. Summary of methods, categories, datasets, and key remarks from recent studies on 3D and 2D anomaly detection approaches in industrial applications.

Author	Year	Category	Dataset	Method	Remarks	Refs
A. Costanzino et al.	2024	3D	MVTec 3D-AD	Representation	Multimodal industrial anomaly detection by cross-modal feature mapping.	[22]
H. Zhu et al.	2024	3D	MVTec 3D-AD	Representation	Group-level feature contrastive learning for high-res 3D anomaly detection.	[9]
Y. Tu et al.	2024	3D	MVTec 3D-AD	Representation	Self-supervised feature adaptation to enhance RGB and 3D feature consistency.	[11]
Y. Cao et al.	2024	3D	MVTec 3D-AD	Representation	Combines 2D and 3D features for improved anomaly detection performance.	[6]
V. Zavrtanik et al.	2024	3D	MVTec 3D-AD	Representation	Depth simulation enhances 3D surface anomaly detection.	[23]
J. Wang et al.	2024	3D	MVTec 3D-AD	Representation	Incremental template neighborhood matching for anomaly detection.	[10]
C. Bi et al.	2024	3D	MVTec 3D-AD	Reconstruction	Lightweight dual-branch reconstruction network for anomaly detection.	[24]
Z. Li et al.	2024	3D	MVTec 3D-AD	Reconstruction	Dual reconstruction network for 3D anomaly detection.	[14]
Z. Gu et al.	2024	3D	MVTec 3D-AD	Knowledge Distillation	Multi-modal reverse distillation integrates RGB, depth, and surface normals.	[15]
T. Batzner et al.	2024	2D	MVTec AD	Knowledge Distillation	Accurate visual anomaly detection at millisecond-level latencies.	[17]
Y. Wang et al.	2023	3D	MVTec 3D-AD	Representation	Hybrid fusion combines multimodal features for robust anomaly detection.	[20]
M. Rudolph et al.	2023	3D	MVTec 3D-AD	Knowledge Distillation	Asymmetric student-teacher networks for industrial anomaly detection.	[16]
P. Bergmann et al.	2023	3D	MVTec 3D-AD	Knowledge Distillation	Adaptation of student-teacher framework for 3D point cloud anomaly detection using deep geometric descriptors.	[25]
Y. M. Chu et al.	2023	3D	MVTec 3D-AD	Representation	Shape-guided dual-memory learning for multi-modal anomaly detection.	[21]
R. Chen et al.	2023	3D	MVTec 3D-AD	Reconstruction	multi-scale and multi-modality feature encoder-decoder combined with a multi-modality anomaly segmentation network.	[26]
E. Horwitz et al.	2023	3D	MVTec 3D-AD	Representation	Classical 3D features are leveraged for 3D anomaly detection.	[19]
S. Bae et al.	2023	2D	MVTec AD	Representation	Position and neighborhood information for industrial anomaly detection.	[27]
H. Jang et al.	2023	2D	MVTec AD	Representation	Neighboring pixel-based industrial anomaly detection.	[7]
J. Lee et al.	2023	2D	MVTec AD	Representation	Memory-efficient visual anomaly detection with incremental dimension reduction.	[8]
L. Hua et al.	2023	2D	MVTec AD	Representation	Self-supervised anomaly detection with pixel-point random walk.	[28]
M. Abdallah et al.	2023	2D	Sensor Signal	Representation	Anomaly Detection and Inter-Sensor Transfer Learning on Smart Manufacturing Datasets.	[29]
K.-J. Jeong et al.	2022	2D	Sensor Signal	Representation	Two-stage deep anomaly detection with heterogeneous signals.	[18]
S. Holly et al.	2022	2D	Industrial Cooling Dataset	Reconstruction	Autoencoder-based anomaly detection with fault localization in cooling systems.	[4]

Taking a different approach, Holly et al. [4] explored fault localization in large-scale industrial cooling systems, where predictive maintenance is complicated by limited labeled data, sensor variability, and complex system structures. Their study highlighted the difficulty of training supervised models

due to the lack of systematically collected failure reports, making unsupervised learning a viable alternative. To address this, they proposed an autoencoder-based anomaly detection workflow designed for multivariate and dynamically evolving time-series sensor data. Their approach incorporated

individual sensor reconstruction errors to localize faults and identify root causes, integrating expert knowledge for improved interpretability. The proposed method demonstrated high consistency in fault localization, offering a practical solution for monitoring and maintaining refrigeration systems in industrial settings.

A particularly noteworthy development within signal-based anomaly detection is the application of Singular Spectrum Analysis (SSA) to rotary encoder signals in industrial robotic systems. Rotary encoders generate high-frequency position data reflecting the precise movements of robot joints and actuators, making them suitable for capturing subtle mechanical irregularities that are difficult to detect visually. Recent works [33], [34], [35] have proposed methods that apply SSA to decompose these time-series signals into interpretable components, allowing the identification of weak positional oscillations and low-amplitude deviations indicative of early-stage faults or mechanical wear.

These approaches often incorporate hierarchical sparsity-driven priors to model the temporal structure of normal operation patterns, improving robustness against background noise and transient fluctuations. Notably, the effectiveness of SSA-based frameworks has been demonstrated in detecting fine-grained anomalies that are functionally equivalent to the small-scale defects targeted in vision-based detection. Although such methods rely on fundamentally different modalities and feature representations compared to image-based techniques, their shared focus on capturing weak deviations from normalcy highlights a complementary perspective. Including these techniques in the broader discussion of industrial anomaly detection enriches the methodological landscape and emphasizes the growing importance of multimodal strategies in handling increasingly complex manufacturing environments.

While prior research on sensor data has significantly contributed to anomaly detection, these studies primarily focused on time-series and sensor signals. However, the increasing complexity of industrial environments has highlighted the need for more comprehensive approaches that integrate multiple data modalities. The introduction of multimodal industrial benchmark datasets such as MVTec 3D-AD [5] has further accelerated progress by providing standardized evaluations for diverse anomaly detection scenarios. These datasets, combined with advances in machine learning techniques, have enabled researchers to develop specialized methods tailored for both 2D and 3D anomaly detection, bridging the gap between traditional sensor-based approaches and multimodal analysis.

As the field progressed, anomaly detection research began to diverge into three key methodologies: representation-based methods, reconstruction-based methods, and knowledge distillation approaches. Each of these methodologies has played a crucial role in enhancing anomaly detection for industrial applications, offering distinct advantages and addressing unique challenges. Representation-based methods focus on learning meaningful and discriminative feature

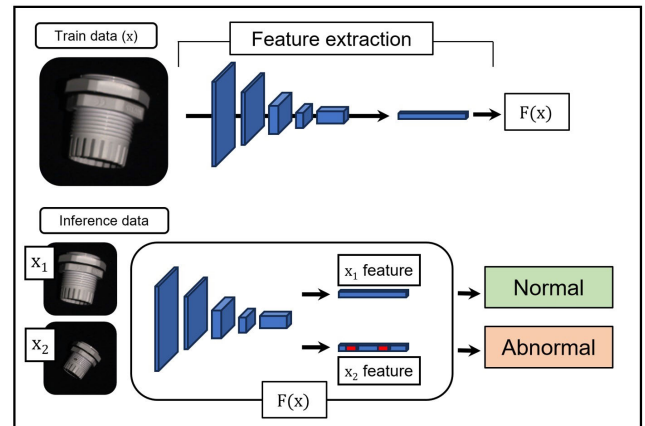


FIGURE 1. The representation-based approach extracts feature representations $F(x)$ from train data. During inference, features extracted from input data are compared to the learned representation to determine whether the sample is normal or abnormal.

representations, enabling robust anomaly identification across various domains. Reconstruction-based methods aim to model normal data distributions and detect anomalies by measuring deviations in the reconstruction process. Meanwhile, knowledge distillation approaches leverage pre-trained models and transfer learned representations to enhance anomaly detection performance, often with reduced computational costs. These methodologies have significantly advanced the field by improving the detection of defects in complex industrial environments.

A. REPRESENTATION-BASED METHODS

Representation-based methods focus on extracting robust and discriminative features from the input data, making them highly effective for identifying anomalies across diverse industrial domains. As illustrated in Fig. 1, these methods extract meaningful feature representations from training data through a feature extraction network, denoted as $F(x)$. During inference, test samples such as x_1 and x_2 undergo the same feature extraction process, generating feature embeddings $F(x_1)$ and $F(x_2)$, respectively. The extracted features are then compared against the learned representations from normal training data. If a given sample closely aligns with normal patterns (e.g., x_1), it is classified as normal. However, if deviations are detected in the feature space (e.g., x_2 , highlighted with red activation differences), the sample is identified as anomalous.

Recent advancements have extended representation-based methods to leverage multimodal inputs, incorporating RGB, depth, and point-cloud data to enhance anomaly detection. For example, Cao et al. [6] proposed a complementary pseudo multimodal feature framework, which integrates 3D local geometric features from handcrafted descriptors with 2D global semantic features from pre-trained neural networks. By projecting 3D point clouds into multi-view 2D images, their method captures rich semantic information, overcoming

the limitations of using only 3D or 2D features. Similarly, Tu et al. [11] enhances RGB-depth anomaly detection by improving cross-modal feature alignment. It introduces intra-modal feature compactness and cross-modal local-to-global consistency to refine feature representations across different modalities, making the detection of small-scale anomalies more effective.

Furthermore, BTF [19] introduced a multimodal framework that combines RGB and 3D shape descriptors to enhance anomaly detection. Their approach, termed back to the feature, integrates handcrafted geometric features with color-based representations, enabling the detection of fine textures and color-based anomalies that might be indistinguishable using only geometric information. This fusion of modalities improves sensitivity to both structural deformations and subtle surface defects, making it well-suited for industrial inspection applications.

Building on this, M3DM [20] proposed memoryless multimodal anomaly detection, which integrates RGB and depth modalities without relying on external memory banks. Their method focuses on reducing computational overhead while maintaining high detection accuracy, addressing the challenge of feature redundancy in multimodal anomaly detection. Employing adaptive feature selection techniques ensures that only the most informative multimodal cues are retained, improving efficiency in industrial inspection settings.

In another effort to enhance multimodal anomaly detection, CFM [22] introduced crossmodal feature mapping, a technique that aligns representations across different modalities to improve anomaly detection. Their method employs feature transformation strategies to ensure consistency between RGB, depth, and point cloud inputs, allowing the detection model to effectively leverage complementary multimodal information. By learning crossmodal relationships, crossmodal feature mapping strengthens the robustness of feature embeddings, particularly in cases where anomalies exhibit subtle variations across different sensing modalities.

Similarly, the shape-guided method [21] developed shape-guided anomaly detection, a framework that utilizes a dual memory bank to align RGB features with point cloud representations. Unlike conventional 3D anomaly detection methods that process geometric data in isolation, their approach integrates signed distance functions with a learned RGB-3D mapping mechanism, enabling a more structured representation of object surfaces. By maintaining separate memory banks for both modalities, their method captures cross-domain feature correspondences, enabling precise identification of geometric deformations, missing components, and fine-grained defects. This multimodal fusion strategy enhances the precision of anomaly localization, making it particularly effective for industrial inspection tasks where both geometric and appearance-based anomalies must be considered.

Beyond multimodal approaches, some methods focus specifically on optimizing 3D-only features. Zhu et al. [9]

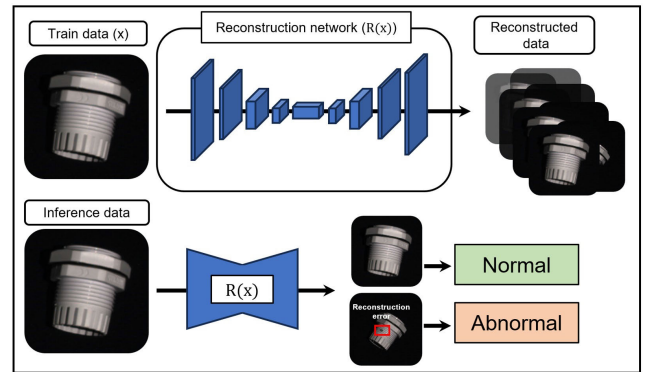


FIGURE 2. The reconstruction-based approach utilizes a reconstruction network $R(x)$ to reconstruct train data. During inference, the network processes the input data and use reconstruction errors to identify anomalies. Normal samples are well reconstructed, whereas abnormal samples exhibit significant reconstruction errors.

introduced a group-level contrastive learning framework to optimize the representation of the feature space in the detection of high-resolution point cloud anomalies. By avoiding downsampling, it preserves fine-grained structural details, while intercluster uniformity networks and intracluster alignment networks improve feature distribution. In addition, adaptive group-center selection prioritizes anomaly-prone regions, enhancing sensitivity in complex industrial datasets. Wang et al. [10] further extends 3D anomaly detection by enabling incremental template-based updates, allowing models to adapt without retraining, which is essential for real-time industrial monitoring.

For 2D anomaly detection, representation-based methods have evolved to focus on spatial feature modeling and efficient feature compression. N-pad model [7] introduced neighboring pixel-based anomaly detection, which refines positional consistency by leveraging neighboring pixel dependencies. Unlike conventional methods that assume perfect image alignment, this approach incorporates Bhattacharyya distance-based weighting to mitigate misalignment errors, improving feature extraction in industrial applications. Lee et al. [8] addressed computational challenges in 2D anomaly detection by introducing a GPU-optimized feature compression technique. Their approach applies incremental singular value decomposition to dynamically reduce feature dimensions, enabling real-time processing with minimal memory overhead.

B. RECONSTRUCTION-BASED METHODS

Reconstruction-based methods have been widely adopted for anomaly detection, particularly for industrial applications where normal data are readily available but anomalous data are scarce. As illustrated in Fig. 2, these methods utilize a reconstruction network, denoted as $R(x)$, which learns to reconstruct normal training data. During inference, test samples are fed into the trained reconstruction model to generate reconstructed outputs. If a given test sample closely

resembles its reconstruction (indicating a low reconstruction error), it is classified as normal. Conversely, if the reconstructed output significantly deviates from the input image (highlighted with red regions indicating high reconstruction error), the sample is identified as anomalous [30]. This approach is particularly useful for detecting structural or texture anomalies by analyzing the discrepancies between the input and the reconstructed representations.

To enhance the effectiveness of 2D reconstruction-based methods, various techniques have been proposed. Liu et al. [12] introduced an edge-enhanced reconstruction approach, where images are reconstructed from edge information rather than raw pixel values. This method forces the network to generate missing color and texture information while preserving high-frequency details, making it less likely to reconstruct anomalies accurately. Additionally, a denoising autoencoder mechanism was incorporated to prevent the network from simply copying edge structures, ensuring better generalization to normal patterns. This approach improves the ability to detect subtle surface anomalies while reducing false positives caused by reconstruction artifacts.

Collin and Vleeschouwer [13] explored strategies to enhance reconstruction robustness by introducing corruption-based training. They trained an autoencoder to reconstruct clean images from corrupted inputs, preventing the model from systematically learning to reproduce both normal and defective regions. To further improve image sharpness, skip connections were employed, allowing key structural information to bypass the bottleneck and preserve finer details. Their approach demonstrated how structured corruption techniques can refine reconstruction-based anomaly detection by making models more resilient to noise and better at distinguishing normal and abnormal patterns.

Recent research has extended reconstruction-based techniques to 3D anomaly detection by leveraging dual-modality inputs. Bi et al. [24] proposed a dual reconstruction network designed for RGB-D anomaly detection, where both RGB and depth information are reconstructed to improve defect localization. Unlike previous approaches that relied on pre-trained models and memory banks, their method directly learns to reconstruct depth and color features, reducing computational overhead. Additionally, an entropy-based feature fusion module was introduced to efficiently combine RGB and depth representations, ensuring that depth features contribute meaningfully to anomaly detection rather than acting as noise. This method highlights the importance of feature fusion and multimodal reconstruction in enhancing 3D industrial defect detection.

C. KNOWLEDGE DISTILLATION METHODS

Knowledge distillation is a training paradigm in which a smaller, more efficient model (student) learns to mimic the behavior of a larger, pretrained model (teacher). This is achieved by transferring knowledge through soft probability distributions, feature representations, or intermediate

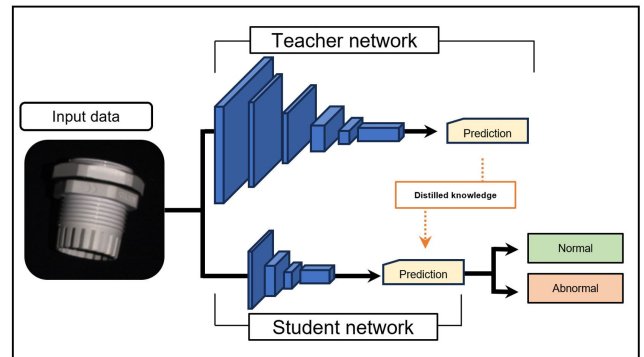


FIGURE 3. The knowledge distillation-based approach employs a teacher-student network framework. The teacher network generates predictions, which are distilled as soft labels and transferred to the student network. The student network, trained with both the original input and the distilled knowledge, produces final predictions to distinguish normal and abnormal samples.

activations rather than relying solely on hard labels. Originally introduced to compress large networks while preserving their performance, knowledge distillation has been widely adopted for anomaly detection [31].

One of the most prominent applications of knowledge distillation in this field is the student-teacher (ST) network, where the student model attempts to replicate the teacher model outputs under normal conditions. During inference, deviations between the two models indicate anomalies because the student model is unable to generalize unseen defective patterns. As illustrated in Fig. 3, the detection of anomalies based on knowledge distillation consists of a teacher network, a large and well trained model that provides high-quality predictions as the standard for normal data behavior, and a student network, a smaller model trained to mimic teacher outputs using distilled knowledge transfer. During inference, both networks process the same input data, and their predictions are compared. If the student network closely matches the teacher's predictions, the input is classified as normal. However, significant discrepancies indicate that the student network has failed to generalize due to its lack of exposure to defective patterns during training, causing the input to be flagged as anomalous [32].

This method is particularly advantageous in real-time industrial applications, as it enables efficient anomaly detection while maintaining high speed with a lightweight model. By leveraging the discrepancy between the teacher and student models, knowledge distillation provides a computationally efficient yet highly effective approach to identifying defects in manufacturing and other industrial domains. This technique is particularly useful in detecting subtle defects and handling multimodal data, making it an essential tool for industrial anomaly detection.

For example, student-teacher networks have demonstrated exceptional performance for real-time anomaly detection. Batzner et al. [17] proposed an optimized knowledge distillation framework called EfficientAD, which balances anomaly

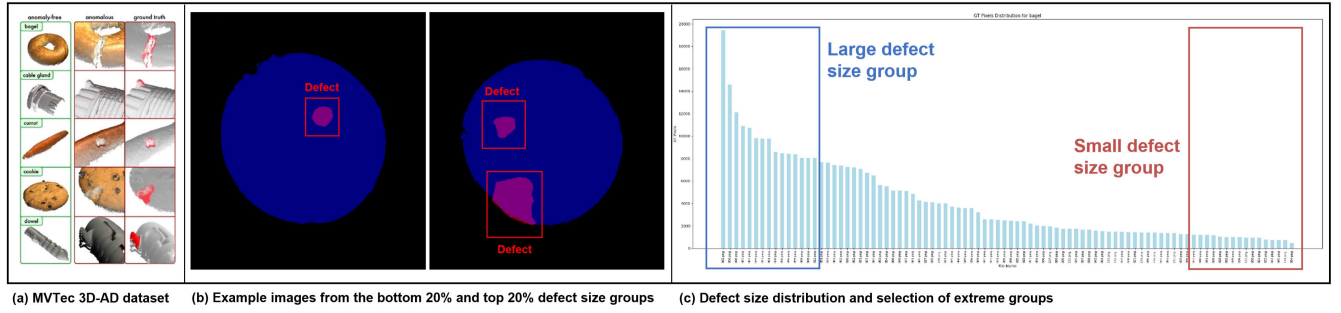


FIGURE 4. Examples of dataset splits used in the experiment. (a) Sample images from the MVTEC 3D-AD dataset, which consists of various industrial objects with both anomaly-free and anomalous samples. The dataset provides ground truth defect masks for precise anomaly localization. (b) Example images from the bottom 20% and top 20% defect size groups. The left image represents a sample from the subset with the smallest 20% defect sizes, while the right image corresponds to a sample from the subset with the largest 20% defect sizes. (c) Distribution of defect sizes within the dataset and the selection of extreme groups. The left (blue) region represents the large defect size group (top 20%), while the right (red) region represents the small defect size group (bottom 20%). These subsets are used to analyze the impact of defect size on anomaly detection performance.

detection performance with computational efficiency. Their method introduces a lightweight student-teacher network capable of millisecond-level inference, addressing the strict latency constraints of industrial applications. To prevent the student from overgeneralizing to anomalous samples, they incorporated loss-induced asymmetry, ensuring that the student fails to reconstruct defects while maintaining strong normal pattern representation. In addition, their model integrates an autoencoder-based logical anomaly detector that detects structural inconsistencies beyond feature-based deviations.

In the 3D domain, Gu et al. [15] introduced a memory-guided knowledge distillation approach to address the “normality forgetting” problem, where student networks struggle to distinguish subtle anomalies from minor variations in normal samples. Their method employs a normality recall memory module, which adaptively retrieves stored normal feature representations to reinforce student feature alignment. This improves the model’s ability to detect fine-grained 3D surface defects while reducing false positives caused by slight deviations in texture or structure.

Similarly, AST [16] proposed an asymmetric student-teacher network, which mitigates the issue of excessive generalization in conventional student-teacher models. Instead of using structurally similar teacher and student networks, they employed a bijective normalizing flow model as the teacher, ensuring that even small variations in input data lead to noticeable changes in feature space. This design enhances the model’s sensitivity to out-of-distribution samples, preventing anomalies from being misclassified as normal. Furthermore, they incorporated positional encoding and foreground masking for 3D images, improving anomaly localization by isolating relevant structural features from background noise.

Despite these advancements, the effects of different defect sizes in multimodal anomaly detection are not well studied. While multimodal approaches have demonstrated improvements in defect detection by integrating RGB, depth, and point-cloud data, most studies focus on optimizing

Algorithm 1 Defect Size-Based Grouping Using Ground-Truth Masks

Input: N : total number of test samples

$rgb_mask[i]$: binary ground-truth defect mask for sample i

Output: $size_score[i]$ for defective samples, and group indices

- 1 Initialize empty list $defect_samples$;
- 2 **for** $i \leftarrow 1$ to N **do**
- 3 $area \leftarrow \sum rgb_mask[i]$; **if** $area > 0$ **then**
- 4 $size_score[i] \leftarrow area$; Append i to $defect_samples$;
- 5 Sort $defect_samples$ by $size_score$ in ascending order;
- 6 Select bottom 20% $\rightarrow$ Small Defect Group;
- 7 Select top 20% $\rightarrow$ Large Defect Group;

overall performance rather than systematically analyzing how these methods perform across different defect sizes. Small-scale defects, such as micro-scratches or minor surface deformations, may require different detection strategies compared to larger structural anomalies, yet few studies have explicitly investigated these variations.

Furthermore, existing multimodal defect detection research primarily evaluates models under general conditions without considering how defect size influences detection performance. Since defect characteristics such as size and shape can significantly impact detection accuracy, understanding the relationship between multimodal configurations and defect scale remains an open challenge. Despite the growing adoption of multimodal sensing in industrial inspection, few studies have systematically examined how different modalities contribute to detecting defects of varying scales.

To address this gap, we analyze the impact of defect size on multimodal anomaly detection performance in the following section. By evaluating detection accuracy across different defect scales, we aim to provide insights into how

multimodal integration affects the robustness of anomaly detection models under varying conditions.

III. EXPERIMENTS

In this section, we evaluate the impact of defect size on anomaly detection performance by comparing five methods in a standardized environment. The experiments assessed both efficiency and accuracy across varying defect scales, where each method utilized the highest-performing model proposed in its corresponding paper for inference experiments. The five selected methods were chosen based on their outstanding defect detection performance in prior multimodal defect detection studies that utilized the MVTec 3D-AD dataset, which consists of both point cloud and RGB data.

The size of the defect plays a crucial role in the detection of anomalies because smaller defects often pose greater challenges due to their subtle visual and geometric deviations [2]. Many existing methods struggle to detect minor defects that can be overshadowed by noise or insufficient feature representation. Conversely, larger defects are more easily distinguishable but may lead to overconfidence in the detection models, potentially reducing generalizability. Addressing these variations is critical for developing robust industrial anomaly detection systems. As illustrated in Figs. 4 (b) and (c), the MVTec 3D-AD dataset exhibits substantial variation in defect sizes, with some defects covering large regions while others remain highly localized. To quantify defect size, we computed the number of pixels labeled as defects in the ground-truth RGB mask and the number of 3D points located within the projected defect area in the point cloud. The final defect size score was defined as a normalized, weighted sum of these two measures. Based on this score, the top 20% and bottom 20% of samples were categorized as large-defect and small-defect groups, respectively. The detailed procedure for computing the defect size and forming the groups is summarized in Algorithm 1. We conducted further evaluations on the two groups separately.

To investigate this, we conducted experiments on the MVTec 3D-AD dataset and analyzed the performance of different methods while varying defect scales. By systematically evaluating the detection accuracy and computational efficiency across defect sizes, we aimed to provide insights into the strengths and limitations of each approach.

A. EXPERIMENTAL SETTINGS

In this study, we utilized the MVTec 3D-AD dataset, which is specifically designed for industrial quality inspection tasks. As shown in Fig. 4 (a), this dataset provides a comprehensive set of 3D point-cloud scans and RGB images for 10 different object classes, such as bagels, carrots, cookies, and foam. It includes a variety of data types: 3-channel TIFF images that store the XYZ coordinates, 3-channel PNG images for the RGB color information of each point, and sensor data containing the internal parameters of the 3D sensor. This

TABLE 2. Comparison of 3D anomaly detection methods in terms of model complexity and inference speed. The table presents the number of parameters, frames per second (FPS), and inference time for each method.

Model	Parameters (M)	FPS	Inference Time (sec)
Shape-guided [21]	14.166	0.650	1.538
M3DM [20]	123.474	0.994	1.012
BTF [19]	69.677	6.630	0.151
CFM [22]	5.535	9.880	0.101
AST [16]	84.673	83.132	0.012

multimodal information is crucial for the accurate detection of defects in real-world industrial settings.

The MVTec 3D-AD dataset consists of 4,147 high-resolution 3D scans divided into 2,656 training samples, 294 validation samples, and 1,197 test samples. Notably, the training set contains only defect-free samples, aligning with the common practice in unsupervised anomaly detection, where the model learns only from normal data. However, the test set includes both defective and nondefective samples, accompanied by ground truth data for evaluation purposes.

To further analyze the impact of defect size on anomaly detection performance, we conducted inferences using three different subsets of the test data: (1) the full test dataset, (2) the top 20% of samples with the largest defects, and (3) the bottom 20% of samples with the smallest defects. This approach enables us to systematically assess which models maintain robust performance regardless of defect size. By comparing the results across these subsets, we aim to identify models that exhibit consistent accuracy and efficiency across varying defect scales and derive insights into strategies for mitigating accuracy degradation, such as feature enhancement techniques or multiscale learning, to ensure reliable anomaly detection in diverse industrial scenarios.

The performance of the models was assessed on the basis of two key aspects: accuracy and efficiency. For accuracy, we employed both image-level and pixel-level evaluation metrics. The former was quantified using the image-level area under the receiver operating characteristic curve (I-AUROC). At the same time, the latter was measured by the area under the per-region overlap curve (AUPRO). Efficiency was assessed in terms of model parameters, frames per second (FPS), and inference time in seconds. These metrics provide a comprehensive assessment of the practical applicability of each model in industrial contexts, ensuring a balance between detection accuracy and the ability to meet real-time processing demands.

The I-AUROC is used to evaluate the performance of an anomaly detection model by summarizing the trade-off between the true positive rate (TPR) and false positive rate (FPR) across various threshold settings. The formula for AUROC is

$$AUROC = \int_0^1 TPR(t) d(FPR(t)) \quad (1)$$

where t denotes the threshold. In practice, the AUROC is calculated by plotting the receiver operating characteristic (ROC) curve, which is a graph of TPR against FPR for different thresholds, and then measuring the area under this curve. For image-level anomaly detection, I-AUROC measures the ability of the model to distinguish between normal and anomalous images across all thresholds, providing an overall performance score ranging from 0 to 1, where 1 indicates perfect detection and 0.5 represents a random guess.

AUPRO evaluates the localization accuracy of detected anomalies at the pixel level. It is computed by integrating the per-region overlap (PRO) scores over a range of false positive rates, providing a fine-grained assessment of how well the predicted anomalous regions align with ground-truth components across varying confidence thresholds. Formally, AUPRO is defined as

$$\text{AUPRO} = \frac{1}{\tau} \int_0^{\tau} \text{PRO}(\text{FPR}) d\text{FPR}, \quad (2)$$

where τ denotes the upper limit of the integration range, typically set to 0.3 or 0.01. In our experiments, we set $\tau = 0.3$. This metric provides a robust evaluation of spatial alignment between predicted anomaly scores and ground-truth masks, complementing image-level metrics such as I-AUROC.

The FPS quantifies the inference speed of the model, specifically in real-time applications such as video or image processing. It is defined as:

$$\text{FPS} = \frac{N}{T} \quad (3)$$

where N is the number of frames (or images) processed and T is the total time taken (in seconds) to process the frames.

A higher FPS value indicates that the model can process more frames per second, demonstrating its efficiency and suitability for real-time anomaly detection tasks in industrial settings.

B. EXPERIMENTAL RESULT

Table 3 presents the anomaly detection performance of each model across the entire test set, encompassing defects of various sizes. The shape-guided method achieves the highest overall scores in both I-AUROC and AUPRO, with values of 0.946 and 0.976, respectively. These results indicate that the model not only excels at detecting anomalies but also demonstrates strong capability in localizing them with pixel-level accuracy. CFM follows with 0.943 in I-AUROC and 0.966 in AUPRO, showing competitive but slightly lower performance. Although models such as AST and BTF exhibit relatively low model complexity and fast inference times, their AUPRO scores, measured at 0.944 and 0.960, respectively, suggest limitations in localization accuracy. This gap implies that, even when anomaly presence is detected correctly, precise spatial alignment with ground-truth masks is not sufficiently achieved by these models. The consistently high AUPRO of the shape-guided model underscores the

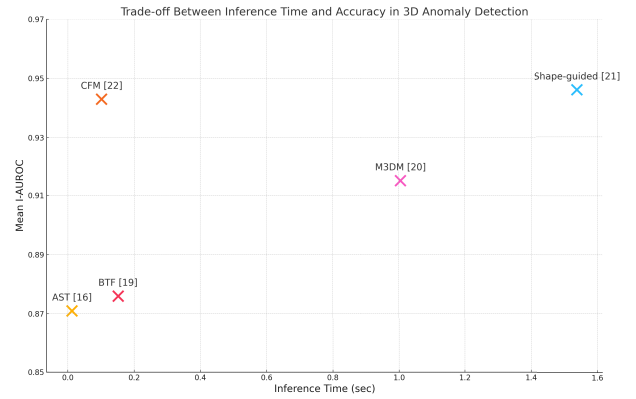


FIGURE 5. Visualization of the trade-off between inference time and detection performance (Mean I-AUROC) across 3D anomaly detection models.

advantage of preserving spatial detail through shape-aware representations.

In the top 20 percent of the largest defects, summarized in Table 4, the shape-guided model maintains a high I-AUROC of 0.979, which matches the best-performing models in detection capability. However, its AUPRO decreases to 0.908, which is lower than the values reported by CFM, M3DM, and BTF, recorded at 0.961, 0.962, and 0.971, respectively. This suggests that while the shape-guided model effectively identifies the presence of large defects, its localization accuracy tends to degrade when dealing with spatially extensive regions. In contrast, models such as BTF and CFM demonstrate better localization precision for large anomalies, likely due to their design choices that favor smooth and generalized spatial features. The AST model, although achieving rapid inference, exhibits significant degradation with an I-AUROC of 0.516 and AUPRO of 0.930, indicating limited reliability in detecting and localizing larger-scale anomalies.

In the bottom 20 percent of the smallest defects, as shown in Table 5, the shape-guided model again achieves the highest I-AUROC, reaching 0.950, which indicates strong detection sensitivity to small-scale anomalies. Nevertheless, its AUPRO declines to 0.854, the lowest among the evaluated models. This result reveals that, although the model effectively detects the presence of small anomalies, it struggles to localize their narrow and irregular boundaries accurately. In contrast, CFM and M3DM report AUPRO scores of 0.966 and 0.964, respectively, indicating superior spatial coherence in predicting small and subtle defect regions. The BTF model also maintains consistent localization performance, with an AUPRO of 0.962 and a moderate I-AUROC of 0.763, demonstrating a balanced approach in both detection and localization across small defect categories.

In summary, the shape-guided model achieves the most stable performance in terms of detection accuracy across all defect sizes, as reflected in the I-AUROC values. However, the observed fluctuations in AUPRO scores across defect scales reveal a limitation in spatial localization under extreme

TABLE 3. Comparison of anomaly detection performance for “All Defect Sizes” using the MVtec 3D-AD dataset. both I-AUROC and AUPRO metrics are reported.

Metric	Method	Mean	Bagel	Cable Gland	Carrot	Cookie	Dowell	Foam	Peach	Potato	Rope	Tire	Avg. Rank
I-AUROC	AST [16]	0.871	0.938	0.953	0.829	0.972	0.974	0.949	0.820	0.532	0.997	0.740	3.2
	BTF [19]	0.876	0.927	0.758	0.967	0.926	0.928	0.623	0.918	0.918	0.914	0.885	4.1
	M3DM [20]	0.915	0.985	0.809	0.970	0.923	0.929	0.893	0.932	0.899	0.944	0.866	3.4
	CFM [22]	0.943	0.989	0.885	0.981	0.990	0.974	0.889	0.956	0.919	0.970	0.874	2.0
	Shape-guided [21]	0.946	0.983	0.892	0.985	0.979	0.976	0.862	1.000	0.966	0.962	0.857	2.2
AUPRO	AST [16]	0.944	0.967	0.943	0.976	0.913	0.904	0.922	0.978	0.981	0.926	0.930	4.4
	BTF [19]	0.960	0.976	0.968	0.979	0.974	0.929	0.900	0.975	0.981	0.949	0.971	3.4
	M3DM [20]	0.960	0.967	0.969	0.978	0.937	0.931	0.930	0.976	0.968	0.971	0.971	3.4
	CFM [22]	0.966	0.977	0.955	0.982	0.941	0.946	0.958	0.981	0.982	0.966	0.974	2.2
	Shape-guided [21]	0.976	0.981	0.972	0.982	0.972	0.961	0.978	0.981	0.983	0.974	0.975	1.1

TABLE 4. Comparison of anomaly detection performance for the top 20% largest defects using the MVtec 3D-AD dataset. Both I-AUROC and AUPRO metrics are reported.

Metric	Method	Mean	Bagel	Cable Gland	Carrot	Cookie	Dowell	Foam	Peach	Potato	Rope	Tire	Avg. Rank
I-AUROC	AST [16]	0.516	0.468	0.429	0.486	0.746	0.515	0.634	0.408	0.424	0.620	0.454	5.0
	BTF [19]	0.942	1.000	0.804	1.000	0.962	0.969	0.766	0.963	0.992	1.000	0.967	2.6
	M3DM [20]	0.958	1.000	0.782	0.986	0.971	0.952	0.975	0.982	0.997	1.000	0.936	2.8
	CFM [22]	0.979	1.000	0.821	0.996	0.998	1.000	1.000	0.991	1.000	1.000	0.981	1.4
	Shape-guided [21]	0.979	0.997	0.978	1.000	0.987	1.000	0.912	1.000	0.977	1.000	0.941	2.1
AUPRO	AST [16]	0.930	0.931	0.927	0.970	0.857	0.861	0.903	0.965	0.977	0.970	0.939	4.2
	BTF [19]	0.971	0.976	0.977	0.982	0.973	0.951	0.931	0.975	0.983	0.982	0.981	1.4
	M3DM [20]	0.962	0.963	0.971	0.978	0.928	0.928	0.930	0.980	0.978	0.981	0.978	2.4
	CFM [22]	0.961	0.973	0.961	0.983	0.913	0.925	0.965	0.982	0.983	0.955	0.971	2.3
	Shape-guided [21]	0.908	0.940	0.852	0.961	0.856	0.824	0.814	0.945	0.959	0.978	0.947	4.6

TABLE 5. Comparison of anomaly detection performance for the bottom 20% smallest defects using the MVtec 3D-AD dataset. Both I-AUROC and AUPRO metrics are reported.

Metric	Method	Mean	Bagel	Cable Gland	Carrot	Cookie	Dowell	Foam	Peach	Potato	Rope	Tire	Avg. Rank
I-AUROC	AST [16]	0.493	0.433	0.571	0.321	0.600	0.502	0.572	0.484	0.333	0.567	0.544	4.9
	BTF [19]	0.763	0.824	0.588	0.870	0.868	0.829	0.534	0.872	0.750	0.620	0.871	3.6
	M3DM [20]	0.848	0.976	0.784	0.952	0.975	0.915	0.653	0.892	0.745	0.760	0.833	2.7
	CFM [22]	0.896	0.965	0.866	0.957	0.989	0.975	0.922	0.866	0.773	0.894	0.751	2.3
	Shape-guided [21]	0.950	0.971	0.975	0.994	0.948	1.000	0.862	0.993	0.972	0.933	0.852	1.5
AUPRO	AST [16]	0.949	0.978	0.951	0.978	0.958	0.940	0.915	0.982	0.983	0.899	0.905	3
	BTF [19]	0.962	0.982	0.977	0.983	0.979	0.956	0.881	0.974	0.977	0.934	0.975	2.2
	M3DM [20]	0.964	0.967	0.978	0.981	0.947	0.937	0.967	0.982	0.940	0.968	0.969	2.5
	CFM [22]	0.966	0.978	0.958	0.982	0.973	0.939	0.957	0.978	0.978	0.965	0.950	2.4
	Shape-guided [21]	0.854	0.868	0.876	0.945	0.705	0.533	0.921	0.924	0.929	0.893	0.946	4.7

conditions, such as very small or very large anomalies. On the other hand, models such as CFM and M3DM provide robust pixel-level accuracy, particularly in challenging scenarios involving fine-grained defects. These findings emphasize the importance of evaluating both detection sensitivity and localization precision when designing anomaly detection systems for industrial applications.

To investigate the relationship between inference speed and anomaly detection performance, we visualized the Mean I-AUROC of each model against its inference time in Fig. 5. The plot reveals a clear trade-off: models such as Shape-guided and CFM achieved high detection performance but required substantially longer inference times. In contrast, AST demonstrated the fastest inference with a relatively lower accuracy. These observations suggest that architectural choices have a significant impact on the balance between speed and accuracy and that no single model can excel in both simultaneously. Fig. 5 helps identify Pareto-efficient models depending on the requirements for the deployment of industrial settings.

While AST shows notable advantages in inference speed, its inferior detection performance for both large and small defects raises questions about its design. We interpret this limitation as stemming primarily from the representational constraints of its student network. AST employs an asymmetric teacher-student structure, with a Normalizing Flow-based teacher and a lightweight CNN-based student. Although this design mitigates the over-regularization often observed in symmetric teacher-student models, it results in a student network with limited feature depth and expressiveness. Consequently, the student struggles to reproduce fine-grained anomaly patterns, particularly in cases of small or low-contrast defects, resulting in reduced anomaly scores and lower detection accuracy. In addition, the Normalizing Flow-based teacher may introduce instability in density estimation, which can affect performance in some instances. However, we consider this a secondary factor, with the primary limitation being the insufficient depth and capacity of the student network. These findings highlight the importance of striking a balance between model compactness and sufficient feature representation when designing fast yet accurate anomaly detection systems.

IV. DISCUSSION

The results demonstrate that the shape-guided method achieves superior and stable anomaly detection performance across all defect scales, a robustness that is not observed in other methods. While existing approaches exhibit noticeable performance degradation for either small or large defects, this trade-off seems to be achieved by sacrificing performance on smaller defects in favor of faster inference. The likely cause of this degradation is the reliance on general loss functions, which do not sufficiently emphasize sensitivity to fine-grained anomalies, unlike specialized loss functions designed to detect small defects. In contrast, shape-guided methods maintains high detection accuracy regardless of

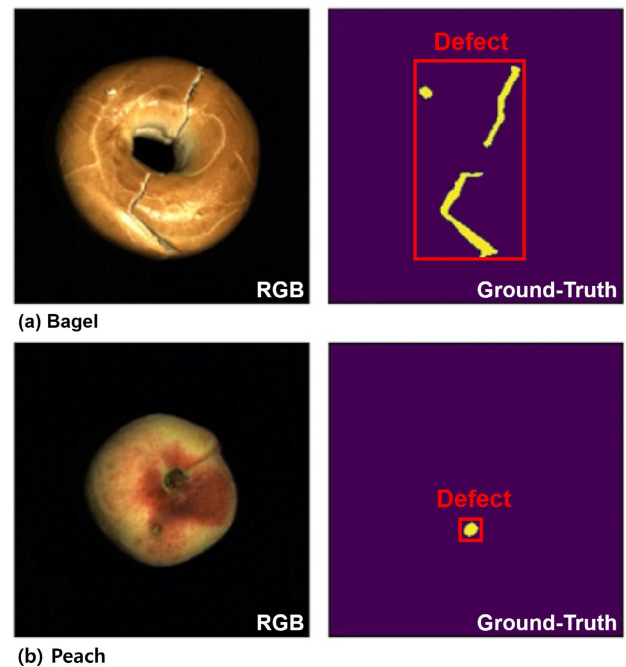


FIGURE 6. Visualization of defect size variations in the dataset. The number of pixels corresponding to the smallest and largest defects varies significantly across different object classes. As shown in (a), the Bagel class exhibits a large defect spanning a significant portion of the surface, with a total of 19,422 pixels, while the smallest defect in this class consists of only 453 pixels. Similarly, in (b), the Peach class contains defects ranging from as few as 350 pixels to as many as 3,654 pixels. The Ground-Truth (GT) annotations illustrate these variations, with larger defects covering extensive regions and smaller defects being more localized. This substantial variation highlights the challenge of maintaining consistent anomaly detection performance across different defect scales.

defect size; however, this advantage comes at the cost of computational efficiency, with shape-guided methods having the slowest inference time among all evaluated methods.

As seen in Figs. 6(a) and (b), the size of the defect varies significantly between different classes of objects, ranging from a few hundred to tens of thousands of pixels. Despite the clear impact of defect scale on detection performance, little prior research has systematically analyzed this factor in anomaly detection. Our study addresses this gap by explicitly investigating how models perform across different defect sizes, providing insights into their robustness and limitations under varying conditions. This analysis highlights the need for designing models that can maintain strong performance across a wide range of defect scales.

Given the excellent performance of shape-guided methods across various defect scales, it would be a promising research direction to improve the inference speed of the shape guidance methods. If its computational efficiency can be optimized while maintaining high detection accuracy, shape guidance could become a more practical solution for real-world industrial applications requiring both accuracy and speed. Future work could explore architectural optimizations such as knowledge distillation, model pruning, or efficient

feature compression to accelerate inference while preserving robust detection capabilities.

By improving shape guidance efficiency, we could develop an anomaly detection framework that not only provides consistent and reliable performance on various defect scales but also meets the demands of high-speed industrial inspection systems. This would bridge the gap between accuracy and efficiency, making high-performance anomaly detection systems more feasible for deployment in real-world manufacturing environments.

By contrast, models that achieve faster inference speed often sacrifice performance for smaller defects. This is likely due to their reliance on standard loss functions, which may not be sensitive enough to capture the fine-grained details of minor anomalies. For the cases where detecting small defects is critical while maintaining fast inference, further research is needed to refine the loss function and model design to balance these competing requirements.

V. CONCLUSION

This study investigated the impact of defect size on anomaly detection performance and inference efficiency. By analyzing multiple state-of-the-art models on the MVTec 3D-AD dataset, we found that defect size plays a critical role in detection accuracy, with certain models exhibiting significant performance drops for either large or small defects. Our findings confirm that balancing accuracy across different defect scales remains a key challenge in anomaly detection. While models such as BTF and M3DM showed strong performance on average, their accuracy noticeably degraded when handling small defects. AST demonstrated rapid inference, but its precision declined sharply at both ends of the defect-size spectrum.

Among the evaluated models, the shape-guided method demonstrated the most stable and highest accuracy in terms of I-AUROC across all defect scales. It also showed competitive performance in AUPRO, indicating strong capabilities in both detection and localization. Notably, it achieved the highest I-AUROC for both the top 20% largest and bottom 20% smallest defect groups. In contrast, AST, despite its efficiency, showed lower localization performance for small and large anomalies. These results highlight that methods solely optimized for speed or global features tend to underperform in cases of subtle or irregular anomalies, especially when pixel-level discrimination is essential.

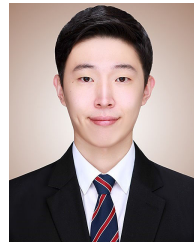
These findings suggest that future research should focus on developing models that optimize both detection robustness and inference efficiency. In particular, the shape-guided model exhibits the slowest processing time per image due to its high-dimensional feature computation, limiting its applicability in real-time settings. On the other hand, knowledge-distillation-based methods such as AST offer significantly faster inference, albeit with less robust performance across defect scales. We can explore techniques such as knowledge distillation, efficient feature compression, and model pruning

to achieve this balance. By addressing these trade-offs, future advancements can contribute to more reliable and practical industrial quality inspection systems.

REFERENCES

- [1] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, "Deep industrial image anomaly detection: A survey," *Mach. Intell. Res.*, vol. 21, no. 1, pp. 104–135, Feb. 2024.
- [2] B. Rolih, D. Ameln, A. Vaidya, and S. Akcay, "Divide and conquer: High-resolution industrial anomaly detection via memory efficient tiled ensemble," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 3866–3875.
- [3] L. Jiang, H. Xu, J. Liu, X. Shen, S. Lu, and Z. Shi, "Anomaly detection of industrial multi-sensor signals based on enhanced spatiotemporal features," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8465–8477, Mar. 2022, doi: [10.1007/s00521-022-07101-y](https://doi.org/10.1007/s00521-022-07101-y).
- [4] S. Holly, R. Heel, D. Katic, L. Schoeffl, A. Stifter, P. Holzner, T. Kaufmann, B. Haslhofer, D. Schall, C. Heitzinger, and J. Kemnitz, "Autoencoder based anomaly detection and explained fault localization in industrial cooling systems," in *Proc. PHM Soc. Eur. Conf.*, Jan. 2022, vol. 7, no. 1, pp. 200–210, doi: [10.36001/phme.2022.v7i1.3349](https://doi.org/10.36001/phme.2022.v7i1.3349).
- [5] P. Bergmann, X. Jin, D. Sattlegger, and C. Steger, "The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization," 2021, *arXiv:2112.09045*.
- [6] Y. Cao, X. Xu, and W. Shen, "Complementary pseudo multimodal feature for point cloud anomaly detection," *Pattern Recognit.*, vol. 156, Dec. 2024, Art. no. 110761.
- [7] J. K. Jang, E. Hwang, and S. H. Park, "N-Pad: Neighboring pixel-based industrial anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Vancouver, BC, Canada, Jun. 2023, pp. 15–24.
- [8] T.-Y. Lee, Y. Nagai, and A. Minezawa, "Memory-efficient and GPU-oriented visual anomaly detection with incremental dimension reduction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Vancouver, BC, Canada, Jun. 2023, pp. 1–9.
- [9] H. Zhu, G. Xie, C. Hou, T. Dai, C. Gao, J. Wang, and L. Shen, "Towards high-resolution 3D anomaly detection via group-level feature contrastive learning," in *Proc. 32nd ACM Int. Conf. Multimedia*, Oct. 2024, pp. 4680–4689.
- [10] J. Wang, X. Wang, R. Hao, H. Yin, B. Huang, X. Xu, and J. Liu, "Incremental template neighborhood matching for 3D anomaly detection," *Neurocomputing*, vol. 581, May 2024, Art. no. 127483.
- [11] Y. Tu, B. Zhang, L. Liu, Y. Li, J. Zhang, Y. Wang, S. Wang, and C. Zhao, "Self-supervised feature adaptation for 3D industrial anomaly detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Oct. 2024, pp. 75–91.
- [12] T. Liu, B. Li, Z. Zhao, D. Xiao, B. Jiang, and L. Geng, "Reconstruction from edge image combined with color and gradient difference for industrial surface anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, LA, USA, Jan. 2022, pp. 1–16.
- [13] A.-S. Collin and C. De Vleeschouwer, "Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Milan, Italy, Jan. 2021, pp. 7915–7922.
- [14] Z. Li, Y. Ge, X. Wang, and L. Meng, "3D industrial anomaly detection via dual reconstruction network," *Int. J. Speech Technol.*, vol. 54, no. 20, pp. 9956–9970, Oct. 2024.
- [15] Z. Gu, L. Liu, X. Chen, R. Yi, J. Zhang, Y. Wang, C. Wang, A. Shu, G. Jiang, and L. Ma, "Remembering normality: Memory-guided knowledge distillation for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, Oct. 2023, pp. 16401–16409.
- [16] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Asymmetric student-teacher networks for industrial anomaly detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2023, pp. 2591–2601.
- [17] K. Batzner, L. Heckler, and R. König, "EfficientAD: Accurate visual anomaly detection at millisecond-level latencies," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2024, pp. 1234–1243.

- [18] K.-J. Jeong, J.-D. Park, K. Hwang, S.-L. Kim, and W.-Y. Shin, "Two-stage deep anomaly detection with heterogeneous time series data," *IEEE Access*, vol. 10, pp. 13704–13714, 2022, doi: 10.1109/ACCESS.2022.3147188.
- [19] E. Horwitz and Y. Hoshen, "Back to the feature: Classical 3D features are (Almost) all you need for 3D anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Vancouver, BC, Canada, Jun. 2023, p. 2968.
- [20] Y. Wang, J. Peng, J. Zhang, R. Yi, Y. Wang, and C. Wang, "Multimodal industrial anomaly detection via hybrid fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, Jun. 2023, pp. 8032–8041.
- [21] Y. M. Chu, L. Chieh, T. I. Hsieh, H. T. Chen, and T. L. Liu, "Shape-guided dual-memory learning for 3D anomaly detection," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Honolulu, HI, USA, Jul. 2023, pp. 6185–6194.
- [22] A. Costanzino, P. Z. Ramirez, G. Lisanti, and L. Di Stefano, "Multimodal industrial anomaly detection by crossmodal feature mapping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2024, pp. 17234–17243.
- [23] V. Zavrtanik, M. Kristan, and D. Skocaj, "Cheating depth: Enhancing 3D surface anomaly detection via depth simulation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2024, pp. 2153–2161.
- [24] C. Bi, Y. Li, and H. Luo, "Dual-branch reconstruction network for industrial anomaly detection with RGB-D data," in *Proc. Int. Conf. Image, Signal Process., Pattern Recognit. (ISPP)*, Guangzhou, China, Jun. 2024, p. 7.
- [25] P. Bergmann and D. Sattlegger, "Anomaly detection in 3D point clouds using deep geometric descriptors," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2023, pp. 2612–2622.
- [26] R. Chen, G. Xie, J. Liu, J. Wang, Z. Luo, J. Wang, and F. Zheng, "EasyNet: An easy network for 3D industrial anomaly detection," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 7038–7046.
- [27] J. Bae, J.-H. Lee, and S. Kim, "PNI: Industrial anomaly detection using position and neighborhood information," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, Oct. 2023, pp. 20818–20827.
- [28] L. Hua, Q. Qi, and J. Long, "P2 random walk: Self-supervised anomaly detection with pixel-point random walk," *Complex Intell. Syst.*, vol. 10, no. 2, pp. 2541–2555, Dec. 2023.
- [29] M. Abdallah, B.-G. Joung, W. J. Lee, C. Mousoulis, N. Raghunathan, A. Shakouri, J. W. Sutherland, and S. Bagchi, "Anomaly detection and inter-sensor transfer learning on smart manufacturing datasets," *Sensors*, vol. 23, no. 1, p. 486, Jan. 2023.
- [30] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009, doi: 10.1145/1541880.1541882.
- [31] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14902–14912.
- [32] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 4182–4191.
- [33] R. N. A. Algburi, H. Gao, and Z. Al-Huda, "Improvement of an industrial robotic flaw detection system," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3953–3967, Oct. 2022.
- [34] R. N. A. Algburi, H. Gao, and Z. Al-Huda, "Implementation of singular spectrum analysis in industrial robot to detect weak position fluctuations," *Fluctuation Noise Lett.*, vol. 20, no. 3, Jun. 2021, Art. no. 2150010.
- [35] R. N. Ali Algburi and H. Gao, "Detecting feeble position oscillations from rotary encoder signal in an industrial robot via singular spectrum analysis," *IET Sci., Meas. Technol.*, vol. 14, no. 5, pp. 600–609, Jul. 2020.



A-SEONG MOON received the M.S. degree from the Department of Artificial Intelligence, Chung-Ang University, Seoul, South Korea, in 2023. He is currently pursuing the Ph.D. degree with the Department of Artificial Intelligence with Chung-Ang University. His research interests include multimodal learning, 3D industrial anomaly detection, and lightweight model development.



KUNHO LEE is currently pursuing the M.S. degree with the Department of Artificial Intelligence, Chung-Ang University, Seoul, South Korea. His research interests include developing multimodal defect detection models for manufacturing. He is particularly focused on optimizing inference speed for industrial anomaly detection systems and improving feature extraction techniques for robust defect identification.



JAESUNG LEE received the B.S., M.S., and Ph.D. degrees in computer science from Chung-Ang University, Seoul, Republic of Korea, in 2007, 2009, and 2013, respectively. He also studies classification, feature selection, and especially multilabel learning with information theory. He is currently the Head and an Associate Professor with the Department of Artificial Intelligence, Chung-Ang University. Concurrently, he is the Chief of the AI/ML Innovation Research Center with Chung-Ang University. His research interests include machine learning, multilabel learning, model selection, and neural architecture search.

...