SUPPLEMENTARY

Anonymous Authors

1 REAL3D-AD DATASET

Real3D-AD, comprising 1,254 high-resolution 3D items with each ranging from forty thousand to millions of points, stands as the most extensive dataset for high-precision 3D industrial anomaly detection currently available. Real3D-AD exceeds other available datasets for 3D anomaly detection in terms of point cloud resolution (0.0010mm-0.0015mm), comprehensive 360-degree coverage, and flawless prototype quality. To ensure both reproducibility and accessibility, the Real3D-AD dataset, benchmark source code, and Reg3D-AD are made available on the website at: https://github.com/M-3LAB/Real3D-AD.

The Real3D-AD dataset includes statistical data across 12 diverse categories. Each category's training set contains only four samples, reflecting a few-shot learning approach akin to 2D anomaly detection. Categories span a variety of toy manufacturing lines such as Airplane, Candybar, and Diamond, among others. The training samples offer comprehensive models of 3D objects, whereas the test samples are only scanned from one side. The dataset's characteristics, including a low ratio of anomaly points and a focus on transparency, make it particularly challenging yet suitable for point cloud anomaly detection tasks.

2 EVALUATION METRICS FOR GROUP3AD

In anomaly detection, particularly in the specialized domain of high-resolution 3D anomaly detection, the precision in evaluating a model's effectiveness is crucial for determining its utility in real-world applications. The Group3AD framework, specifically engineered for this purpose, incorporates a suite of advanced metrics tailored to accurately gauge the model's proficiency in detecting and pinpointing anomalies within detailed 3D environments. This section provides a foundational overview of the evaluation metrics used in Group3AD, detailing their theoretical basis and the significance they hold in enhancing the model's diagnostic capabilities.

These metrics are integral to verifying the robustness and reliability of Group3AD in various industrial and technological contexts, where precise anomaly localization can significantly impact operational safety and efficiency. The selection of these metrics is grounded in their ability to provide a comprehensive assessment of performance across different aspects of anomaly detection, from general detection accuracy to specific localization precision, ensuring that Group3AD meets the rigorous demands of high-stakes applications.

2.1 O-AUROC (Object-level Area Under the Receiver Operating Characteristic Curve)

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR, sensitivity) against the false positive rate (FPR, 1 - specificity) at various threshold settings. The AUROC represents the area under the ROC curve and provides an aggregate measure of performance across all possible classification thresholds. A model with perfect discrimination has an AUROC of 1.0, indicating it can perfectly differentiate between the classes across all thresholds.

O-AUROC assesses the overall ability of the model to distinguish between normal and anomalous objects as complete entities. It is particularly effective for evaluating performance in datasets with varied anomaly presence, ensuring the model's robustness in practical scenarios.

2.2 P-AUROC (Point-level Area Under the Receiver Operating Characteristic Curve)

Similar to O-AUROC, the P-AUROC focuses on the model's performance at a granular level—each individual point in a 3D point cloud. This metric is crucial for applications requiring detailed analysis within objects, where each point's classification directly impacts the overall utility of the detection system.

P-AUROC provides a measure of the model's precision in classifying each point within an object, essential for tasks requiring fine-grained anomaly localization, such as in manufacturing or structural integrity assessments where small anomalies could signify significant defects.

2.3 O-AUPR (Object-level Area Under the Precision-Recall Curve)

The Precision-Recall (PR) curve shows the trade-off between precision and recall for different thresholds. Unlike the ROC curve, a PR curve provides a more informative picture of an algorithm's performance when the classes are very imbalanced. The area under the PR curve (AUPR) is particularly useful as it gives a single measure of performance that considers both the precision and the recall of the predictive model, making it ideal for evaluating models in skewed datasets.

O-AUPR is crucial for assessing Group3AD's efficacy in contexts where anomalies are rare or very subtle, ensuring the model does not overwhelmingly misclassify normal objects as anomalies—a common issue in highly imbalanced datasets.

2.4 P-AUPR (Point-level Area Under the Precision-Recall Curve)

P-AUPR extends the concept of O-AUPR to the granularity of point-level evaluations within each object. This metric is particularly important when precise anomaly localization is crucial, and where false positives (normal points misclassified as anomalies) could lead to unnecessary actions, such as further inspections or repairs.

Together, these metrics provide a robust framework for evaluating the Group3AD system, ensuring it meets both the sensitivity and specificity requirements essential for practical deployments in industrial settings. These metrics not only assess the effectiveness of anomaly detection and localization but also help in tuning the model to improve its performance across diverse operational scenarios.

3 IMPLEMENTATION DETAILS

Building upon the foundation established by the benchmark method Reg3D-AD, the Group3AD framework introduces significant enhancements to the encoder training process to better adapt to high-resolution 3D anomaly detection tasks. Similar to Reg3D-AD, the memory_size of Group3AD was set to 10000 in the experiment. The other parameters and abnormal score calculation method of Group3AD in the experiment are also consistent with Reg3D-AD to ensure fairness.

IUN&IAN used to reinforce the encoder pre trained with Point-MAE. Taking airplanes as an example. Each airplane point cloud is divided into 4096 groups during the training process, each containing 128 neighboring points. Our method has 524288 available points, far exceeding the 8192 points used in conventional PointMAE pre training. Our method more effectively utilizes the information contained in high-resolution point cloud data, avoiding the waste of anomaly point cloud information in the downsampling process.

AGCS prioritizes areas likely to contain anomalies by analyzing local geometric variations within the point clouds. In experiments, the center points obtained by AGCS occupied 20% of the total center points, indicating that the system focused more intensively on these areas, improving the sensitivity and accuracy of the anomaly detection process. In the training process of IUN&IAN, the original point cloud selects the center point through FPS, without using our proposed AGCS module to obtain a more uniform feature distribution in the memorybank.