

Industrial Anomaly Detection Under Background Clutter: A Foreground Extraction Study with RGB and 3D Data

1st Gibeom Kim

Artificial Intelligence
University of Science & Technology
Daejeon, Korea
chroion@etri.re.kr

2nd Hyejin S. Kim

Artificial Intelligence Robot Research Division
Electronics and Telecommunications Research Institute
Daejeon, Korea
marisan@etri.re.kr

Abstract—Automated quality inspection in real-world factories must contend with complex backgrounds that can obscure subtle product defects. We investigate whether explicitly removing background regions benefits multimodal anomaly detection on RGB + 3D point-cloud data. Focusing on washing-drum assembly, we isolate the drum via 3D spatial filtering to create Foreground-Only inputs and compare them with the unprocessed Original scenes. Three state-of-the-art unsupervised models, Asymmetric Student-Teacher (AST) [2], Shape-Guided Dual-Memory [3], and 3DSR [4], are trained solely on normal samples and evaluated on a balanced test set. Image-level AUROC rises consistently for all models when using foreground data: from 0.973 to 0.991 for 3DSR, 0.983 to 1.000 for AST, and 0.949 to 0.999 for Shape-Guided, yielding a mean gain of approximately 3 percentage and a perfect score for AST. Qualitative inspection shows that background removal eliminates false positives and concentrates anomaly heatmaps on genuine defects. These results demonstrate that foreground extraction is a simple yet powerful preprocessing step for RGB + 3D anomaly detection in cluttered industrial environments and should be considered a standard component of deployment pipelines.

Index Terms—Anomaly detection, Machine learning, Machine vision, Industrial inspection, 3D point clouds, Unsupervised learning

I. INTRODUCTION

Unsupervised anomaly detection has become vital in industrial quality inspection, where only normal data are available for training, and the goal is to detect novel defects or anomalies during operation. [6]–[9] Traditional AD benchmarks like MVTec AD focus on 2D images of products, and newer datasets such as MVTec 3D AD [1] extend this to RGB-D / 3D data for detecting geometric and visual anomalies. In these benchmarks, each object is typically captured in isolation against a simple planar background [2], making it straightforward to segment the object from the background. In practice, however, industrial scenes are often more complex: products may be embedded in cluttered environments with rails, fixtures, and other objects in view [13], [14]. This background clutter can confound anomaly detection models by introducing irrelevant features and variations. Recent studies have suggested that removing background regions allows models to focus on the essential object features, thereby improving

detection performance. In this paper, we investigate the impact of explicit foreground extraction on RGB+3D anomaly detection performance in a real-world industrial setting. We consider an assembly scenario of washing machine drums, where anomalies or defects in the drum must be detected from combined color and 3D data [10], [11]. We compare model performance with and without a preprocessing step that removes background data, using multiple state-of-the-art AD models. Our findings show that focusing on the foreground object yields higher detection accuracy, even reaching perfect detection in one case, highlighting the importance of robust foreground isolation in industrial anomaly detection.

II. RELATED WORKS

A. Industrial Anomaly Detection Datasets

The MVTec Anomaly Detection (MVTec AD) [5] dataset introduced a standardized benchmark for unsupervised anomaly detection on high resolution images of various objects and textures. It contains 15 categories with only normal samples in training and dozens of defect types in the test set. Also, the MVTec 3D AD [1] dataset provides RGB-3D point cloud scans for 10 object categories to facilitate detection of 3D structural anomalies as well as visual defects. In both datasets, the training set consists solely of normal products and the test set includes both normal and anomalous samples, with pixel or point level ground truth annotations for defects. Notably, the objects in MVTec 3D AD are scanned in isolation, which makes it easy to obtain a foreground mask for the object. This is an important detail because it means that models can be evaluated mostly on the object itself without background interference. Many state-of-the-art methods implicitly or explicitly take advantage of this. The general expectation is that focusing on the object should improve anomaly detection by eliminating the need to model background regions that are not relevant to the defect.

B. RGB and 3D Anomaly Detection Methods

With the availability of RGB+3D point cloud sensors, recent works have aimed to combine color and 3D geometry modal-

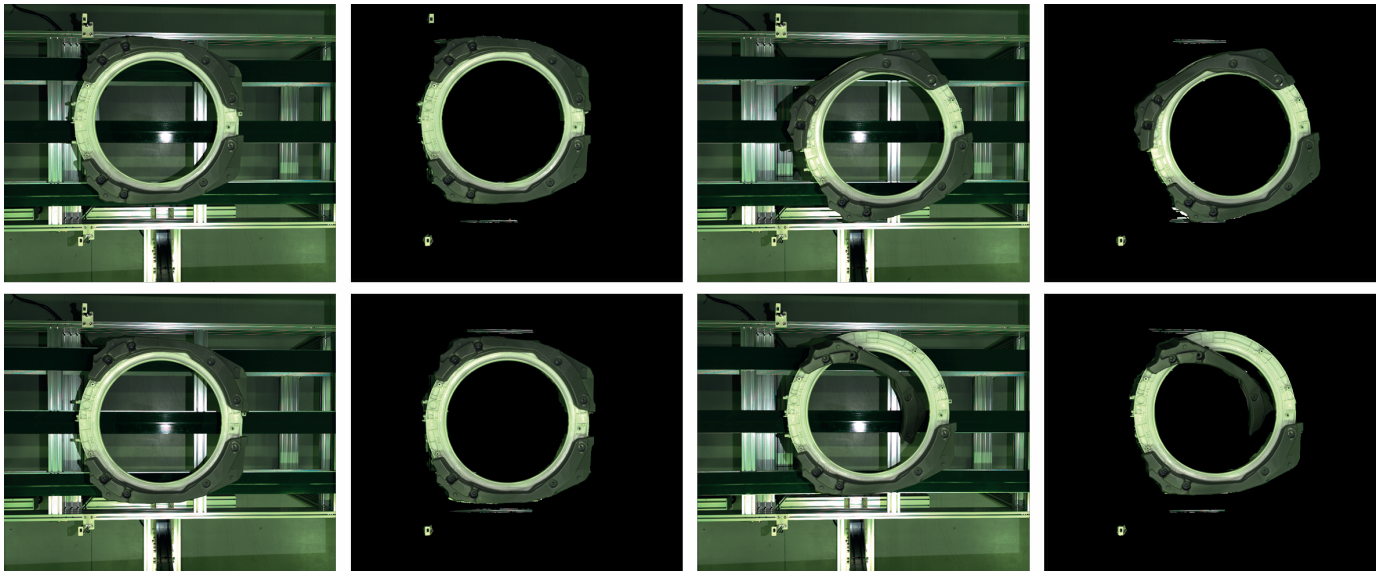


Fig. 1. Example RGB captures from the washingdrum inspection dataset illustrating the effect of 3D foreground extraction. Each pair shows the *Original full scene* and the corresponding *Foreground-Only* image. Columns 1–4 (left block) depict normal drums; Columns 5–8 (right block) depict anomalous drums. Foreground extraction isolates the product surface and removes extensive, highly variable background clutter present in the assembly environment.

ities to detect subtle anomalies. The Asymmetric Student-Teacher (AST) [2] model employs student-teacher networks [12], [14] for anomaly detection. This model’s teacher network produces embeddings for normal data, and a student network is trained to imitate the teacher. Anomalies are detected when the student fails to replicate the teacher’s output. Also this model use a probabilistic normalizing flow as the teacher and a standard feedforward CNN as the student, forcing a large output divergence for anomalous inputs. Another approach is to leverage memory networks or prototype features from normal data named as Shape-Guided Dual-Memory method [3] that trains two expert models: one modeling 3D shape patterns and another modeling 2D appearance, both built from anomaly-free samples. At test time, anomalies are localized by comparing input features against these normal memories in a shape-guided manner. This method achieved top performance on MVTec 3D-AD, benefiting from per-point 3D representations and the fusion of complementary RGB and depth cues. The use of separate memories for geometry and color allows detecting defects that manifest in shape as well as those visible only in color. Recently, a method called 3DSR (3D Surface Anomaly Detection via Depth Simulation) [4] has recently been introduced. Their approach tackles the fact that some surface anomalies are nearly invisible in RGB alone by explicitly enhancing and utilizing depth information. They develop a Depth-Aware Discrete Autoencoder (DADA) to jointly encode RGB and depth data into a shared latent space, and use simulated depth images to augment limited training data. The 3DSR model achieved state-of-the-art accuracy on the challenging MVTec 3D-AD benchmark, outperforming previous methods in both detection accuracy and speed. This highlights the benefit of incorporating geometric depth cues for anomaly detection, as certain defects can be detected from

3D shape even when color appears normal.

III. METHOD

A. Dataset and Preprocessing

Our study is conducted on an industrial anomaly detection task in a washing machine drum assembly line. The data consist of paired RGB images and 3D point clouds captured by an industrial 3D vision sensor observing the assembly process. The goal is to detect abnormal situations or defects that can occur during the drum assembly. Examples of anomalies in this context might include deformed drums, misaligned components, or foreign objects attached to the drum. A key challenge in this real-world dataset is the complex environment: apart from the washing drum (the target object of interest), the scene includes a metal rail and various surrounding structures. This means a substantial portion of each image/pointcloud is background or unrelated objects, unlike the simple backgrounds in MVTec 3D-AD. To address this, we perform a foreground extraction step to isolate the drum. Specifically, we leverage prior knowledge of the drum’s position in the scene to filter out points that belong to the background. All points outside the drum’s region are removed. The result is a cropped point cloud and image containing only the drum. This yields two versions of the dataset: (1) Original (Full) – the raw RGB images and 3D data with background included, and (2) Foreground-Only – the processed data where only the drum and its immediate vicinity remain. Both versions carry the same anomaly labels.

B. Anomaly Detection Models

We evaluate three recent unsupervised anomaly detection models that accept combined RGB and 3D inputs: AST, Shape-Guided, and 3DSR. For each model, we use its published architecture and recommended training procedure, with

TABLE I
ANOMALY DETECTION ACCURACY (AUROC) FOR EACH MODEL, ON ORIGINAL DATA VERSUS FOREGROUND-ONLY DATA. HIGHER IS BETTER.

Model	AUROC (Original)	AUROC (Foreground-Only)	Δ	Rel.%
3DSR [4]	0.9731	0.9910	+0.0179	+1.84
AST [2]	0.9827	1.0000	+0.0173	+1.76
Shape-Guided [3]	0.9490	0.9990	+0.0500	+5.27
Average	0.9683	0.9967	+0.0284	+2.9

the only difference being whether background removal is applied to the input data.

- **AST (Asymmetric Student-Teacher)** – A normalizing-flow teacher and a CNN student are trained on normal data; anomalies are detected from the teacher-student feature discrepancy. A 3D mask can be applied to ignore non-object regions, so removing the background lets the model focus exclusively on the drum.
- **Shape-Guided Dual-Memory** – Maintains two memory banks: one for 3D shape features, one for RGB appearance. At test time it flags points that diverge from either memory. Eliminating background points prevents false matches and sharpens defect localization on the drum surface.
- **3DSR (Depth-Simulation Autoencoder)** – Jointly encodes RGB and depth via a depth-aware discrete autoencoder, augmented by simulated depth maps. With background removed, the network reconstructs only the drum region, boosting sensitivity to subtle surface defects.

IV. RESULTS

I summarize the anomaly detection results for the three models on the washing drum test set, comparing the Original vs Foreground-Only data. We report the image-level AUROC (in fraction form) for each case.

All three models achieve higher AUROC when using only the foreground data, confirming that removing the background leads to better anomaly detection performance. On average, the AUROC across the models improved from about 0.968 (96.8%) with the full scene to 0.997 (99.67%) with just the foreground, an increase of roughly 2.9 percentage points. The Shape-Guided model in particular showed a dramatic jump, from 0.949 to 0.999, indicating that it struggled the most with background clutter and benefited greatly from focusing only on the object. The 3DSR and AST models were already very accurate on the full data (97–98% AUROC), but still saw meaningful improvements to around 99–100% with foreground only input. Notably, the AST model reached an AUROC of 1.000 (100%) when background was removed. It means that model perfectly distinguished all defective samples from all normal samples in our test. This suggests that, at least for this dataset, the combination of AST’s student-teacher framework with a clean foreground input allowed it to detect every anomaly without a single mistake. In contrast, AST’s AUROC was 0.9827 with the original data, implying a few defects were missed or some normal parts of the background were incorrectly flagged as anomalous. Eliminating the background resolved those issues.

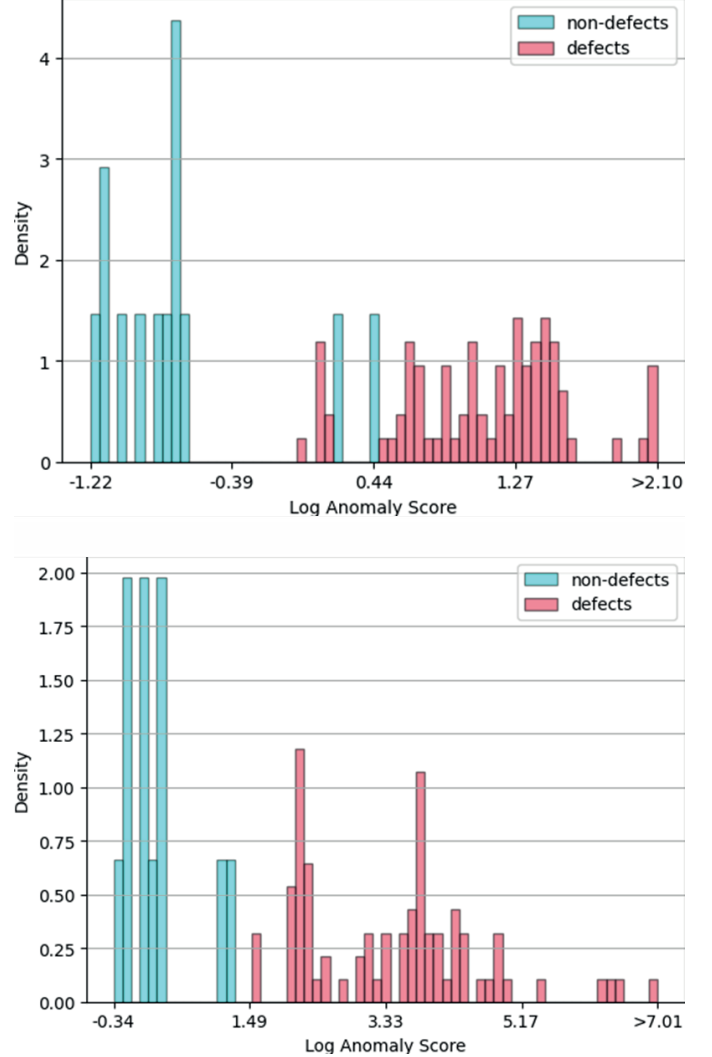


Fig. 2. Distribution of log anomaly scores from the AST model for normal (cyan) and defective (magenta) washing drum samples. **Top:** Model trained and tested on *Original* full-scene data. Substantial overlap between the two score distributions indicates uncertain discrimination and leads to suboptimal image-level accuracy. **Bottom:** Same model configuration trained and tested on *Foreground Only* data obtained via 3D ROI masking. The score distributions separate cleanly, yielding near-perfect decision separability.

V. CONCLUSION

We have presented a comparative study on anomaly detection using RGB and 3D point cloud data in an industrial assembly context, examining the effect of foreground extraction (background removal) on detection performance. Using a washing drum assembly dataset with complex scene

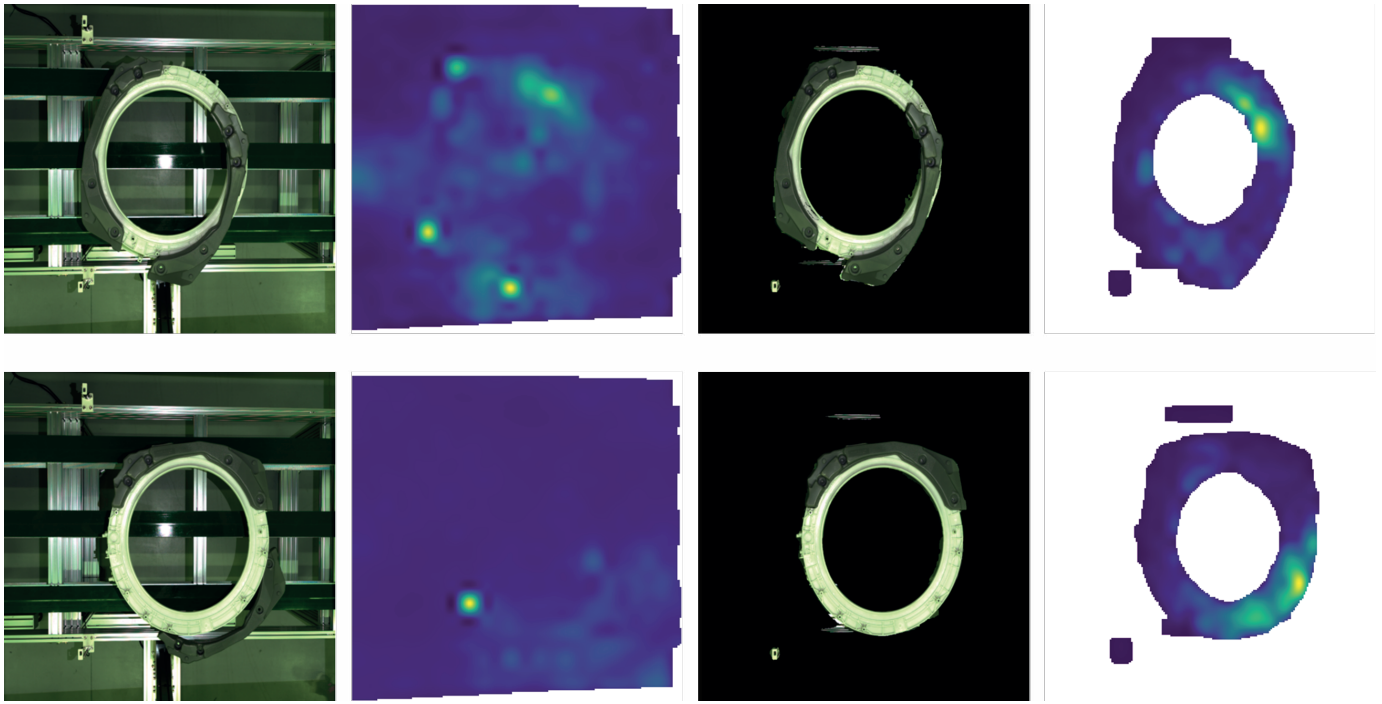


Fig. 3. AST anomaly localization on two defective washing drum samples (rows). Columns: (1) *Original* full-scene RGB input; (2) AST anomaly heatmap for the Original input, showing diffuse and inaccurate prediction, background driven responses; (3) *Foreground Only* RGB obtained by 3D ROI masking; (4) AST anomaly heatmap after foreground extraction, with responses concentrated on true defect regions and minimal background noise. Foreground isolation sharpens localization and reduces false activations on rails and other clutter.

background, we showed that three different state-of-the-art AD models all achieved superior results when the input was limited to the target object, as opposed to the full scene. In particular, isolating the drum and removing background clutter led to an average improvement of around 3% in AUROC, and even enabled one model (AST) to reach a perfect detection score on our test set. These results empirically validate the intuition that background regions can act as distractors or sources of false alarms in anomaly detection. By eliminating these regions, the models can focus exclusively on the product’s appearance and shape, thereby more reliably identifying deviations caused by true defects. This finding has practical significance for deploying anomaly detection in real factories. It suggests that incorporating a foreground segmentation or localization step in the inspection pipeline can substantially boost the accuracy of defect detection. In cases where the object’s location is known or can be obtained, applying a region mask to analyze only the object is highly beneficial. For future work, one could explore automated methods to segment the object from cluttered backgrounds such that the benefits of foreground-focused anomaly detection can be realized without manual intervention. Additionally, while our study focused on image-level detection, an interesting extension would be to evaluate the impact of background removal on pixel-level anomaly localization performance, as one would expect cleaner segmentation of defects when the models are not confused by background noise. In conclusion, our research underscores that a seemingly simple preprocessing step can yield outsized

gains in anomaly detection performance. As advanced models continue to improve and combine multimodal inputs, ensuring that these inputs are free of irrelevant content will remain important. Especially in complex industrial environments, intelligent foreground extraction may be a necessary component of a robust automated visual inspection system. By focusing on products themselves, we can achieve more accurate and trustworthy detection of anomalies, ultimately helping maintain high quality standards in manufacturing processes.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-01049, Development of teaching-less product assembly system for smart factory based on autonomous robot task planning and manipulation)

REFERENCES

- [1] BERGMANN, Paul, et al. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. arXiv preprint arXiv:2112.09045, 2021.
- [2] RUDOLPH, Marco, et al. Asymmetric student-teacher networks for industrial anomaly detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023. p. 2592-2602.
- [3] CHU, Yu-Min, et al. Shape-guided dual-memory learning for 3d anomaly detection. 2023.
- [4] ZAVRTANIK, Vitjan; KRISTAN, Matej; SKOČAJ, Danijel. Cheating depth: Enhancing 3d surface anomaly detection via depth simulation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024. p. 2164-2172.

- [5] BERGMANN, Paul, et al. MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 9592-9600.
- [6] ROTH, Karsten, et al. Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 14318-14328.
- [7] DEFARD, Thomas, et al. Padim: a patch distribution modeling framework for anomaly detection and localization. In: International conference on pattern recognition. Cham: Springer International Publishing, 2021. p. 475-489.
- [8] ZAVRTANIK, Vitjan; KRISTAN, Matej; SKOČAJ, Danijel. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021. p. 8330-8339.
- [9] LI, Chun-Liang, et al. Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 9664-9674.
- [10] LI, Zhuo, et al. 3D Industrial anomaly detection via dual reconstruction network. *Applied Intelligence*, 2024, 54.20: 9956-9970.
- [11] LI, Kecen, et al. DAS3D: dual-modality anomaly synthesis for 3d anomaly detection. In: European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024. p. 148-165.
- [12] BERGMANN, Paul, et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 4183-4192.
- [13] ZHU, Wenbing, et al. Real-IAD D3: A Real-World 2D/Pseudo-3D/3D Dataset for Industrial Anomaly Detection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. 2025. p. 15214-15223.
- [14] CHRISTIANSEN, Peter, et al. DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors*, 2016, 16.11: 1904.