Master’s Thesis Presentation for the 143th Dissertation Evaluation

A Study on Multimodal Industrial Anomaly Detection for Irregular Surfaces using Dynamic Noise Augmentation

동적 노이즈 증강을 이용한 불규칙한 표면의 멀티모달 산업용 이상 탐지 연구

Kunho Lee

dlrmsghz@gmail.com

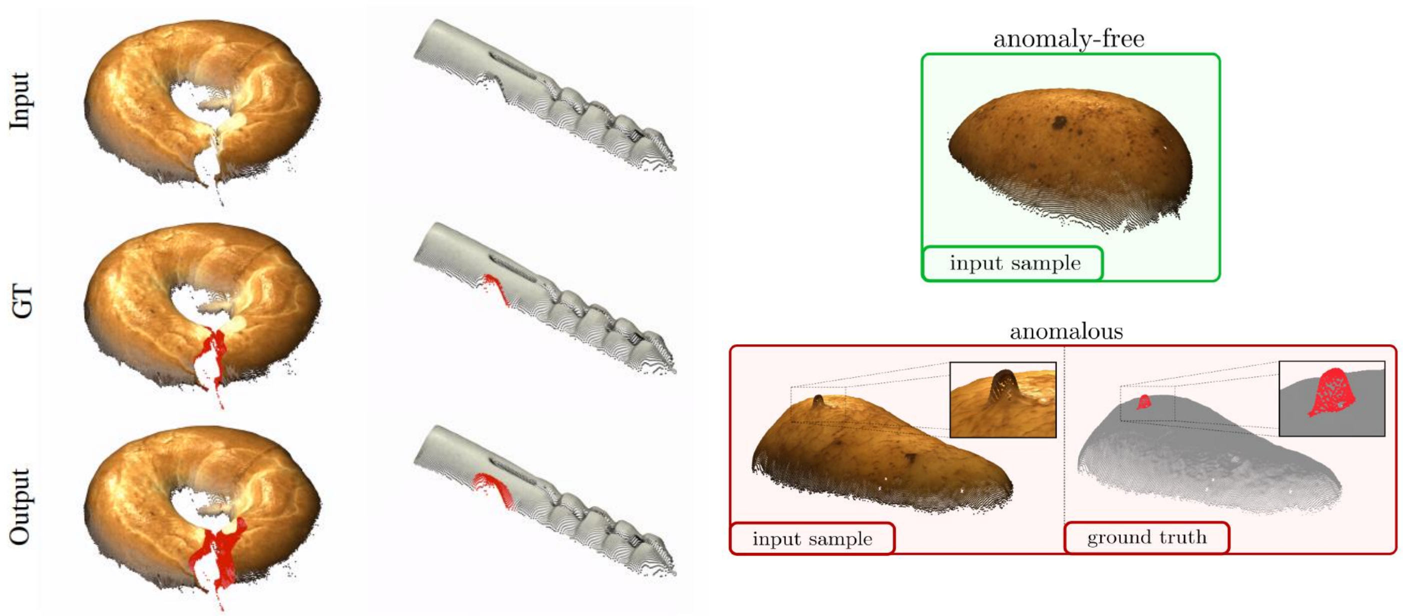
25th Apr, 2025

Contents

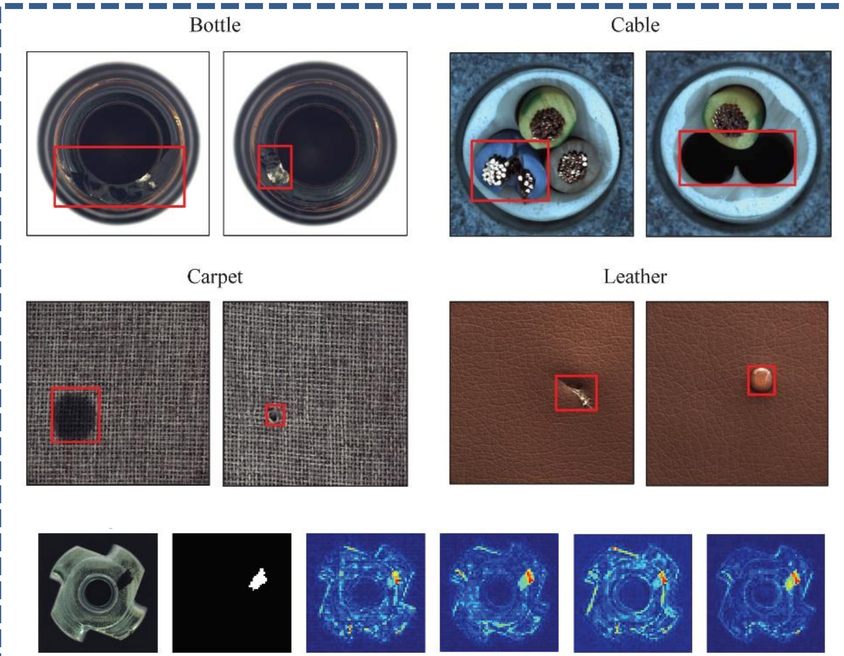
Introduction

Anomaly detection using multimodal data in industrial environments

- In the industrial sector, Anomaly detection is crucial for product quality assurance, ensuring consistently high product quality, and optimizing operational efficiency by identifying deviations from manufacturing and maintaining processes [1].
- Traditional methods have been used for image defect detection, but recent advancements show a shift towards utilizing multimodal data, to enhance overall performance and thereby increase the efficiency of the manufacturing process.



< Figure 1. > Defect detection using 3D point clouds industrial data



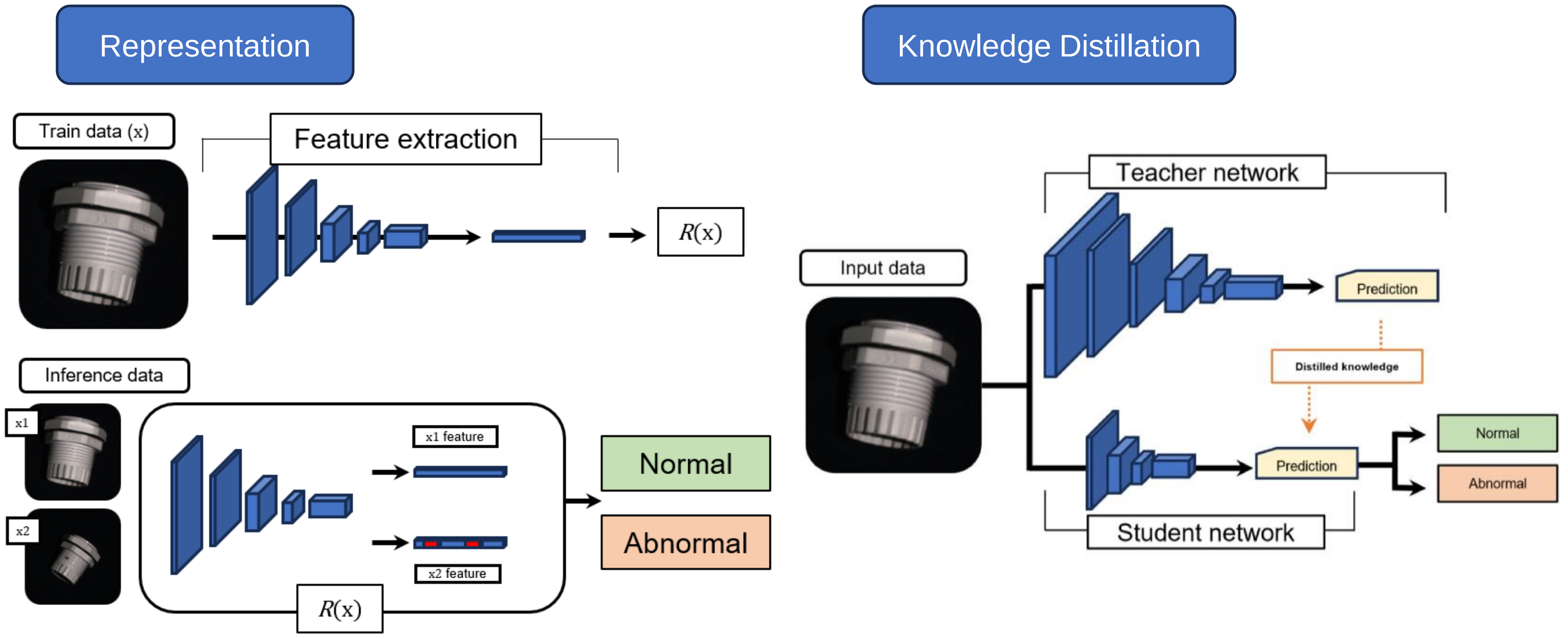
< Figure 2. > 2D Industrial image defect detection

[1] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 14–19, 2020, pp. 4183–4192.

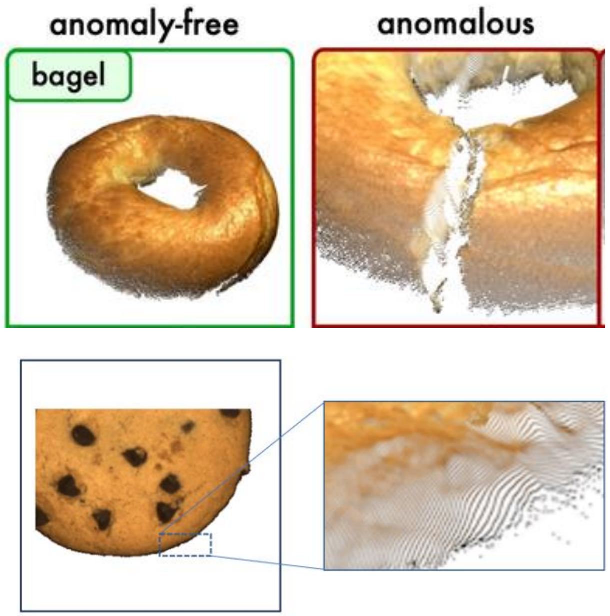
Introduction

Objective and Standard procedure

- The objective of multimodal industrial anomaly detection is to accurately identify various types of defects in industrial environments by leveraging both RGB images and point cloud data, ensuring robust detection across different materials and surface conditions.
- The multimodal defect detection involves extracting feature representations from normal RGB images and point cloud data, comparing them to detect deviations, and ultimately classifying the output as either normal or anomalous conditions [2].



< Figure 3. > Overview of the operation of defect detection using point clouds



< Figure 4. > Real-world 3D point clouds data

[2] Chu, Y.-M., Liu, C., Hsieh, T.-I., Chen, H.-T., & Liu, T.-L. (2023). Shape-guided dual-memory learning for 3D anomaly detection. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, July 23–29, 2023, pp. 6185–6194.

Introduction

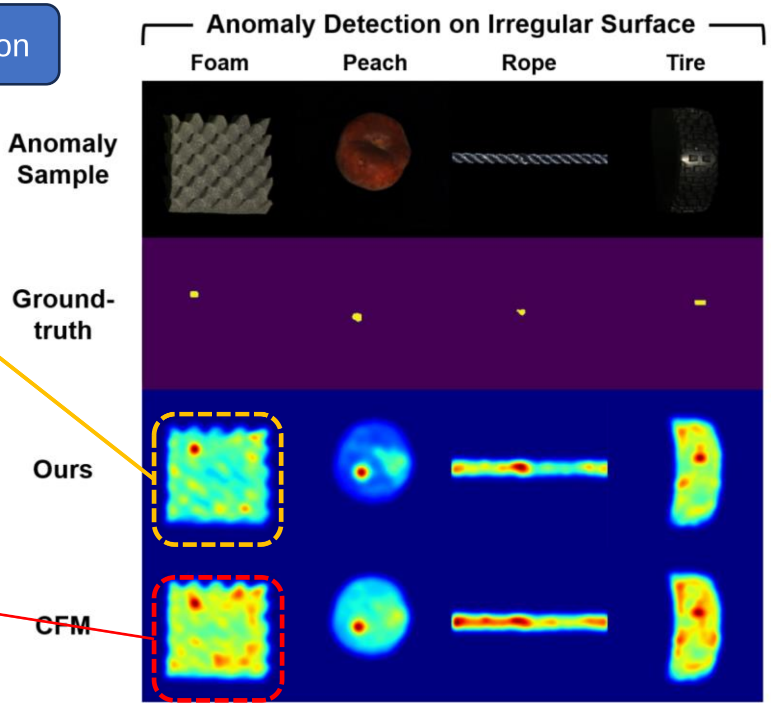
Problem and Strategy

- Irregular industrial surfaces often cause depth distortions and low-intensity regions in 3D point clouds, making it difficult for multimodal anomaly detection to preserve spatial consistency, maintain geometry, and distinguish defects from natural variations.
- This study dynamically enhances weak response regions by injecting learnable Gaussian noise, thereby improving 3D feature representation, preserving fine-grained structural information, and enabling more robust and accurate anomaly localization.

Results of defect detection

Injecting Gaussian noise amplifies feature vectors in low-intensity regions.

Sample with uneven surface geometry



Low-intensity
- Regions with relatively low feature vector norms compared to other areas
- Caused by structural ambiguity, depth distortion, or noise
- These areas are easily overlooked or misinterpreted during training

Noise Injection
- Inject Gaussian noise $N(\mu,\sigma)$ into low-intensity regions
- Enhances feature representation and amplifies weak signals
- Noise parameters μ,σ are learned and optimized during training for maximum performance gain

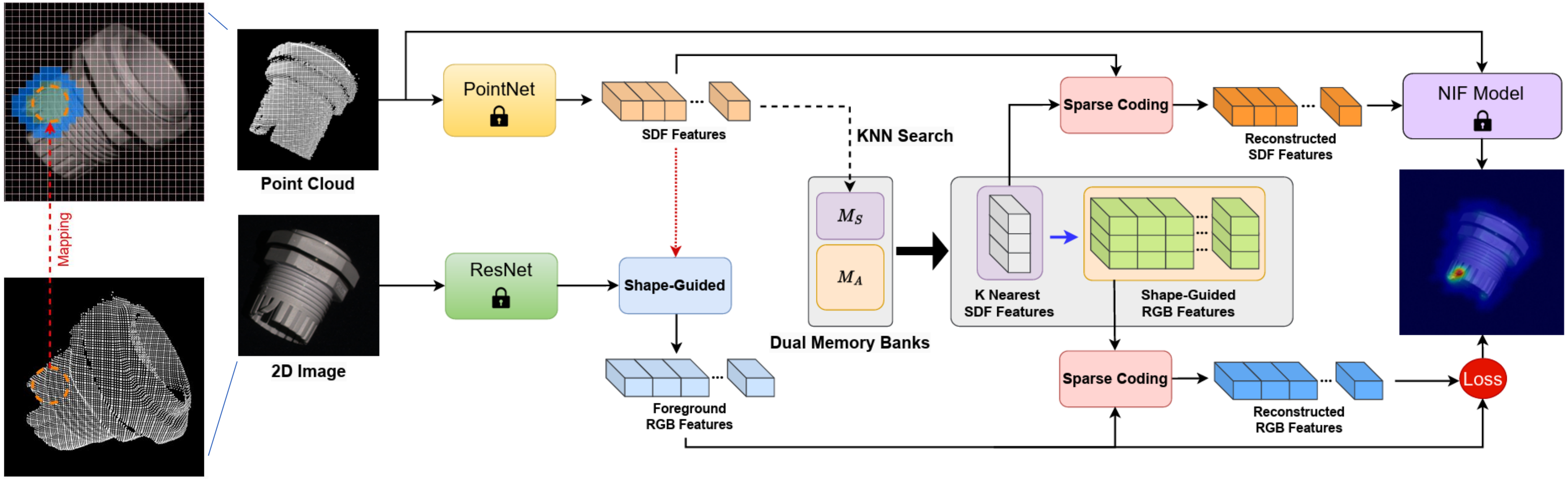
Enhanced feature
- Resulting features are strengthened and more informative
- Repeated training updates noise parameters to maximize anomaly sensitivity Leads to more accurate and robust anomaly detection

< Figure 5. > Qualitative comparison of anomaly detection results on irregular surface data using Dynamic Noise Augmentation

Related Work

Shape-guided dual-memory learning for 3D anomaly detection

- This research focus on 3D anomaly detection, employing two expert models that integrate local geometric information with 2D RGB features to detect structural and color anomalies, improving the accuracy and robustness of defect identification processes [2].
- A drawback of Shape-guided Dual-Memory Learning for 3D Anomaly Detection is the impractical inference time due to the high computational cost of extracting shape features from 3D point clouds and managing dual-memory mechanisms for feature retrieval.



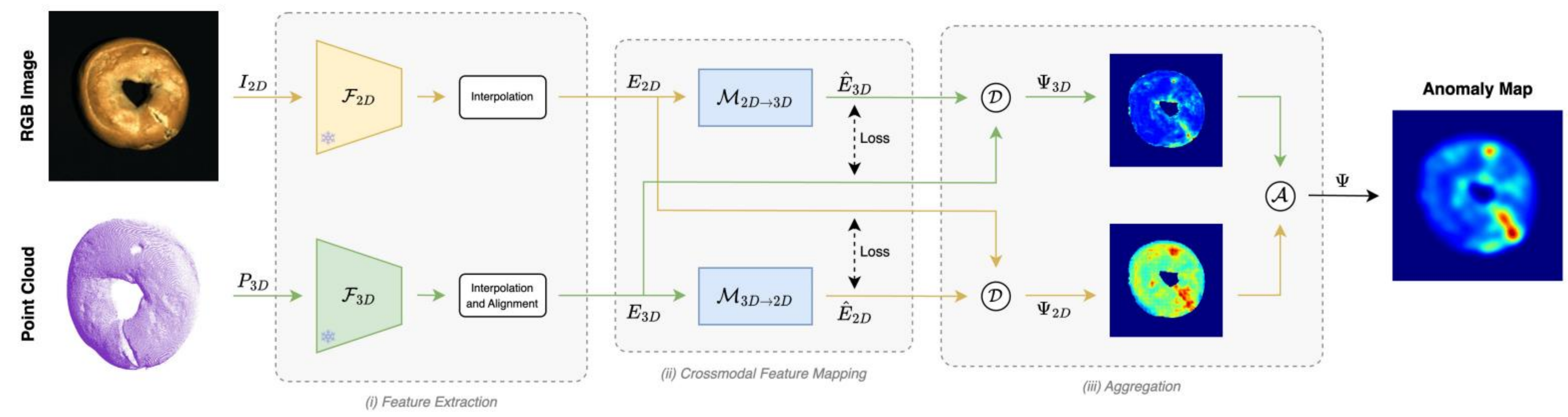
< Figure 6. > Corresponding shape-guided inference using sparse representation

[2] Chu, Y.-M., Liu, C., Hsieh, T.-I., Chen, H.-T., & Liu, T.-L. (2023). Shape-guided dual-memory learning for 3D anomaly detection. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, July 23–29, 2023, pp. 6185–6194.

Related Work

Multimodal industrial anomaly detection by crossmodal feature mapping

- The research proposes a novel multimodal industrial defect detection method that effectively leverages crossmodal feature mapping to seamlessly integrate complementary information from different data modalities for improved defect detection [3].
- While this approach offers fast inference speed and serves as a strong counterpart, it comes at the cost of lower accuracy, particularly when detecting small defects where identifying subtle anomalies becomes more challenging.



< Figure 7. > The overview of Multimodal industrial anomaly detection by crossmodal feature mapping

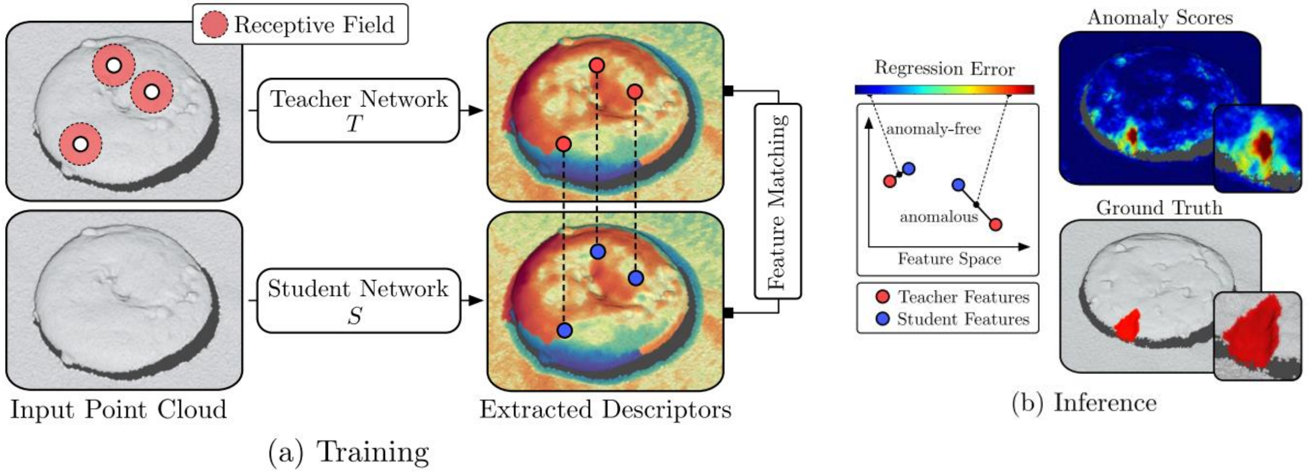
[3] Costanzino, A., Ramirez, P. Z., Lisanti, G., & Di Stefano, L. (2024). Multimodal industrial anomaly detection by crossmodal feature mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 17–21, 2024, pp. 17234–17243.

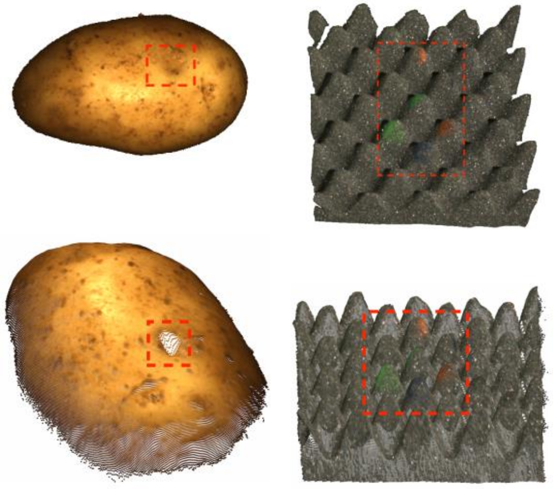
Related Work

Back to the Feature: Classical 3D Features are (Almost) All You Need for 3D Anomaly Detection

- The research addresses the issue of traditional 3D defect detection and segmentation methods underperforming compared to 2D color-based approaches by proposing the approach combining rotation-invariant 3D point cloud descriptors [4].
- Although the fusion of 2D and 3D data aimed to enhance both speed and accuracy, it achieved strong performance in terms of speed but struggled with detection, particularly at the image level performing worse than in other studies.

3D Feature Descriptor





Shape anomalies, such as sharp deformations in objects like cookies and potatoes, are detectable in 3D but not in color views, while texture anomalies, like deformations in color variations in foam, are identifiable in 2D images but not through 3D information.

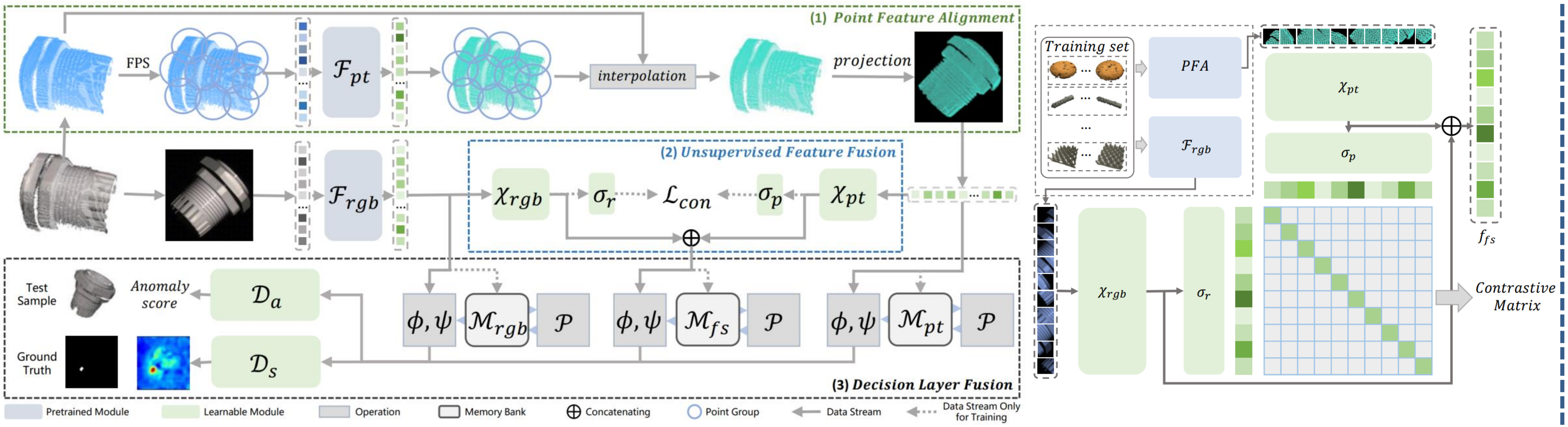
< Figure 8. > Enhancing 3D feature descriptor for anomaly detection with 2D features

[4] Horwitz, E., & Hoshen, Y. (2023). Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, June 18–22, 2023, pp. 12345–12354.

Related Work

Multimodal Industrial Anomaly Detection via Hybrid Fusion

- The research introduces an advanced hybrid fusion approach for multimodal learning that significantly enhances feature interaction and reduces disruptive interference, thereby markedly improving overall accuracy and efficiency [5].
- A drawback of this method is the inefficient inference time due to the computational overhead of managing dual-memory banks performing contrastive learning-based sparse coding, which cause additional processing complexity and resource consumption.

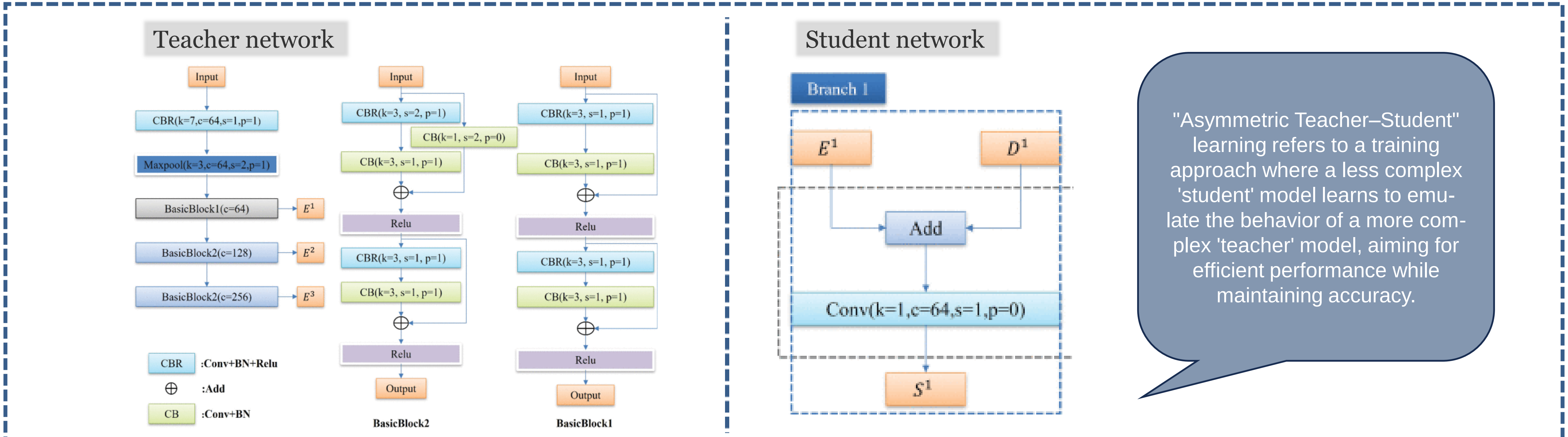


< Figure 9. > The process of Multimodal Industrial Anomaly Detection via Hybrid Fusion

[5] Wang, Y., Peng, J., Zhang, J., Yi, R., Wang, Y., & Wang, C. (2023). Multimodal industrial anomaly detection via hybrid fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, June 18–22, 2023, pp. 23456–23465.

Asymmetric student-teacher networks for industrial anomaly detection

- The research focuses on designing a lightweight and asymmetric knowledge distillation model to achieve real-time defect detection and state-of-the-art performance in both defect detection and localization [6].
- While this method benefits from fast inference due to the lightweight student network, its performance tends to be lower compared to other approaches that directly learn a unified feature representation without relying on teacher-student mappings.



< Figure 10. > The overview of Asymmetric student-teacher networks

[6] Rudolph, M., Wandt, B., & Rosenhahn, B. (2023). Asymmetric student-teacher networks for industrial anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, January 3–8, 2023, pp. 34567–34576.

Related Work

Common Drawback and Solution

- Existing multimodal anomaly detection methods often suffer from either slow inference time due to complex, computation-heavy dual-stream architectures, or poor detection accuracy when optimized solely for speed and deployment.
- To address this trade-off, recent approaches attempt to improve cross-modal feature interaction efficiency or simplify backbone structures, but most still fail to maintain a strong balance between detection performance, robustness, and inference efficiency.

Method	Venue	Summary	Drawback
Shape-guided [2]	ICML, 2023	Multimodal defect detection, employing two expert models that integrate local geometric information with 2D RGB features to detect structural and color anomalies, improving the accuracy and robustness of defect identification processes.	Slow inference time
CFM [3]	CVPR, 2024	A novel multimodal industrial defect detection method that effectively leverages crossmodal feature mapping to seamlessly integrate complementary information from different data modalities for improved defect detection.	Fast inference time / Impractical accuracy
BTF [4]	CVPR, 2023	Addressing the issue of traditional multimodal defect detection and segmentation methods underperforming compared to 2D color-based approaches by proposing the approach combining rotation-invariant 3D point cloud descriptors.	Fast inference time / Impractical accuracy
M3DM. [5]	CVPR, 2023	Hybrid fusion approach for multimodal learning that significantly enhances feature interaction and reduces disruptive interference, thereby markedly improving overall performance.	Impractical accuracy
AST [6]	WACV, 2023	The lightweight and asymmetric knowledge distillation model to achieve real-time defect detection and state-of-the-art performance in both defect detection and localization.	Fast inference time / Impractical accuracy

Proposed Method

Notation and Definition

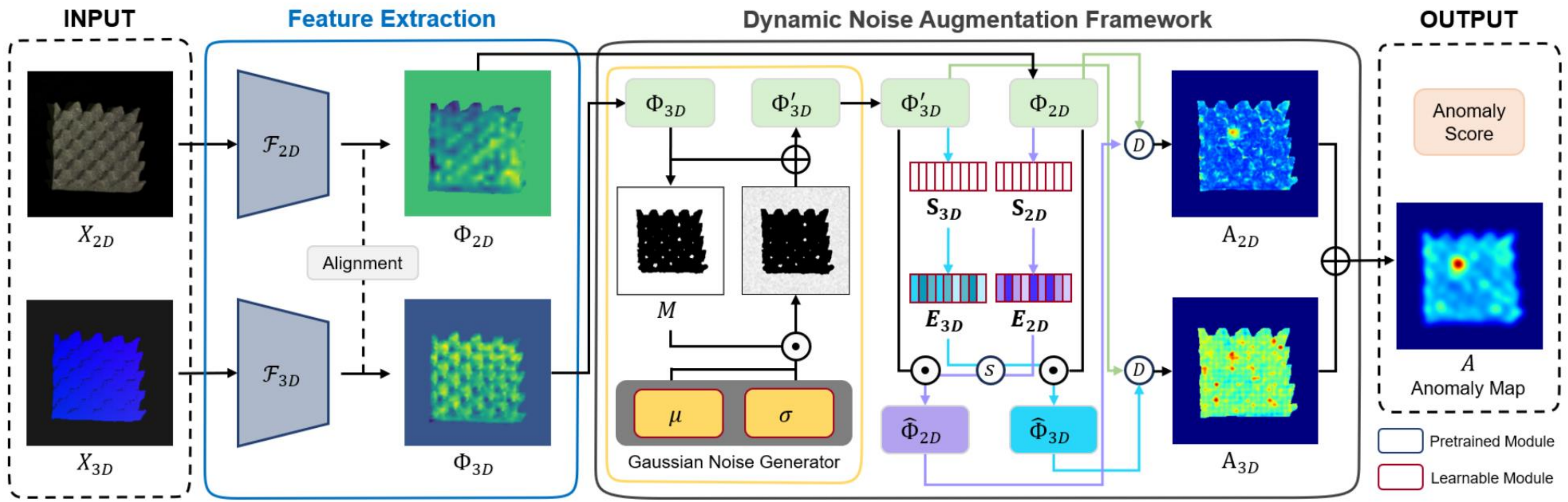
- This study denotes the extracted 3D point cloud features as Φ_{3D} and generate a binary mask M by analyzing feature intensity values to identify weak or ambiguous low-intensity regions that require enhancement during training.
- Dynamic Noise Augmentation (DNA) is a learnable mechanism that selectively enhances weak geometric feature representations by injecting adaptive Gaussian noise $\mathcal{N}(\mu, \sigma^2)$, thereby improving structural fidelity and enabling robust anomaly detection.

Notation	Description	Usage
X_{3D}, X_{2D}	Input 3D point cloud and 2D RGB image data	Modalities used for multimodal feature extraction
F_{2D}	2D feature extraction function	Generates 2D features from RGB image
F_{3D}	3D feature extraction function	Generates 3D features from point cloud
Φ_{3D}, Φ_{2D}	Extracted features from 3D and 2D inputs	Base representations before enhancement
M	Binary mask for low-intensity 3D regions	Selects weak areas for noise injection
μ	Mean of Gaussian distribution	Center of noise injection distribution
Σ^2	Standard deviation of Gaussian distribution	Controls noise intensity/variance
$\mathcal{N}(\mu, \sigma)$	Learnable Gaussian noise distribution	Injected into $\Phi_3{}^D$ to enhance low-intensity features
Φ'_{3D}	Augmented 3D features after noise injection	Refined representation used for cross-modal interaction
$\hat{\Phi}_{2D}, \hat{\Phi}_{3D}$	Final enhanced 2D and 3D features after SE block	Used to generate anomaly scores (A_{2D}, A_{3D})
A_{2D}, A_{3D}	Anomaly maps from 2D and 3D modalities	Localized prediction of abnormal regions
A	Final fused anomaly map	Aggregated output for decision making

Proposed Method

Procedural Overview

- The proposed method extract modality-specific deep features from RGB images and 3D point cloud inputs using pretrained encoders then identify structurally ambiguous low-intensity regions in the 3D space via a binary mask for targeted enhancement.
- Learnable Gaussian noise is dynamically injected into those weak feature areas to strengthen geometric fidelity and the refined features from both modalities are fused and scored to generate a robust and accurate final anomaly map.



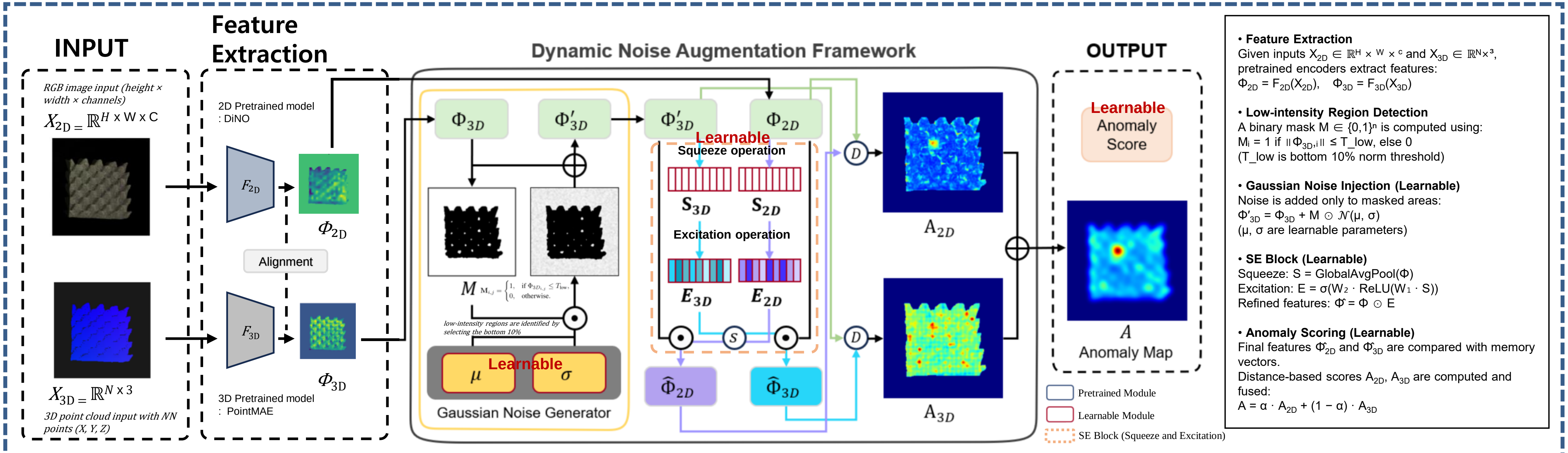
< Figure 11. > Overall architecture of the proposed Dynamic Noise Augmentation (DNA) framework for multimodal anomaly detection

Proposed Method

Details

- A binary mask M is generated by selecting the bottom 10% of 3D feature norms, and learnable Gaussian noise $\mathcal{N}(\mu, \sigma)$ is injected only into these weak regions to produce enhanced features $\Phi'_{3D} = \Phi_{3D} + M \odot \mathcal{N}(\mu, \sigma)$.

- Final features are refined through SE blocks using squeeze and excitation operations, and anomaly scores A_{2D} and A_{3D} are fused as $A = \alpha \cdot A_{2D} + (1 - \alpha) * A_{3D}$, where α is a weighting factor that balances 2D and 3D modalities for robust anomaly detection.

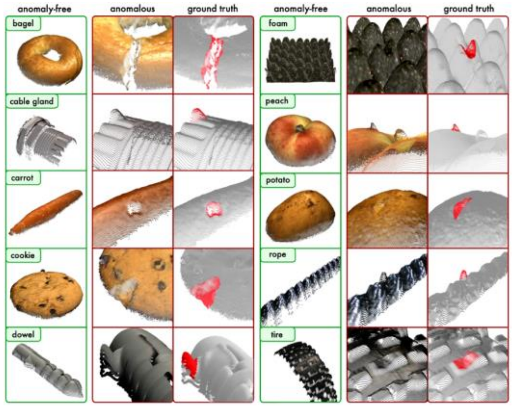


< Figure 12. > Comprehensive overview of the Dynamic Noise Augmentation (DNA) framework with detailed mathematical operations and learnable components.

Experimental Settings

Dataset

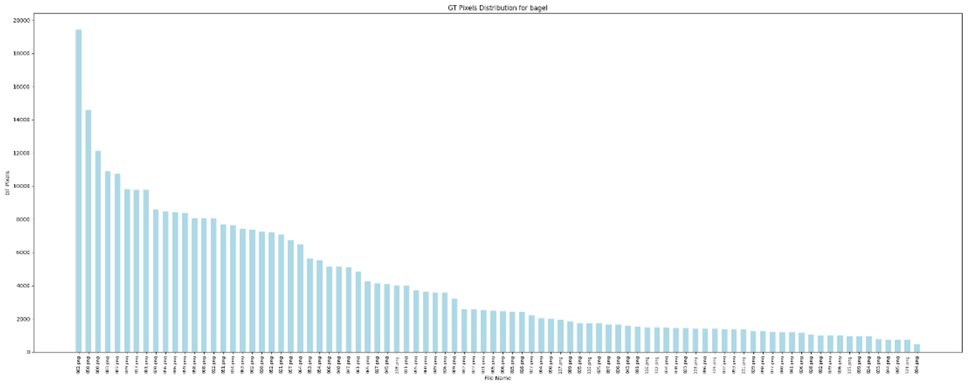
- The MVTec 3D dataset, designed for industrial defect detection, includes 4147 high-resolution 3D scans with depth information across 10 object classes, consisting of 2656 training samples, 294 validation samples, and 1197 test samples.
- The test data was divided based on defect size into the entire dataset, the top 20% largest defects, and the bottom 20% smallest defects, allowing for an evaluation of each method's performance in relation to defect size.



Images data, which contains 3-channel TIFF images that store the x, y and z coordinates.

Images data contains 3-channel PNG images that store the corresponding RGB values for each 3D point.

Sensor data contains the internal camera parameters of the 3D sensor.



< Figure 10. > MVTec 3D-AD, the 3D point clouds dataset for Industrial Quality Inspection

Cross Validation

- The train and valid images (2,952) are split 80:20, excluding the 20% portion of 591 images, and the remaining 2,361 images are divided 75:25 into 1,771 training and 590 validation images, with 1,197 test images.

Experimental Settings

Hyperparameter Settings

- Batch size: 2
- Epochs : 50
- Loss function: Cosine Similarity Loss (Cosine Distance Loss)
- Optimizer : Adam
- Learning Rate: 1e-4

Evaluation Measures

- *Effectivity: Image-level defect detection performance with the area under the receiver operator curve (I-AUROC)*
- *Efficiency: Model parameters, Frames Per Second (FPS), Inference time (sec)*

Experimental Results

Performance Comparison (Inference Speed)

- The proposed method achieves a superior trade-off between accuracy and inference speed, outperforming all comparison models with fewer parameters and high AUC.

Model	Model parameters (M)	FPS	**Mean**	**Inference time (< 1.0) (sec)**
Proposed method	5,768	9.601	0.942 ± 0.004	0.104
AST (2023, WACV) [6]	84,673	83.132	0.842 ± 0.003	0.012
CFM(2024, CVPR) [3]	5,535	10.101	0.929 ± 0.004	0.114
BTF (2023, CVPR) [4]	69,677	6.630	0.863 ± 0.007	0.151
M3DM (2023, CVPR) (Fail) [5]	123,474	0.994	0.915 ± 0.015	1.012
Shape-Guided Dual-Memory Learning (2023, ICML) (Fail) [2]	14,166	0.650	0.945 ± 0.005	1.538

Experimental Results

Performance Comparison (I-AUROC)

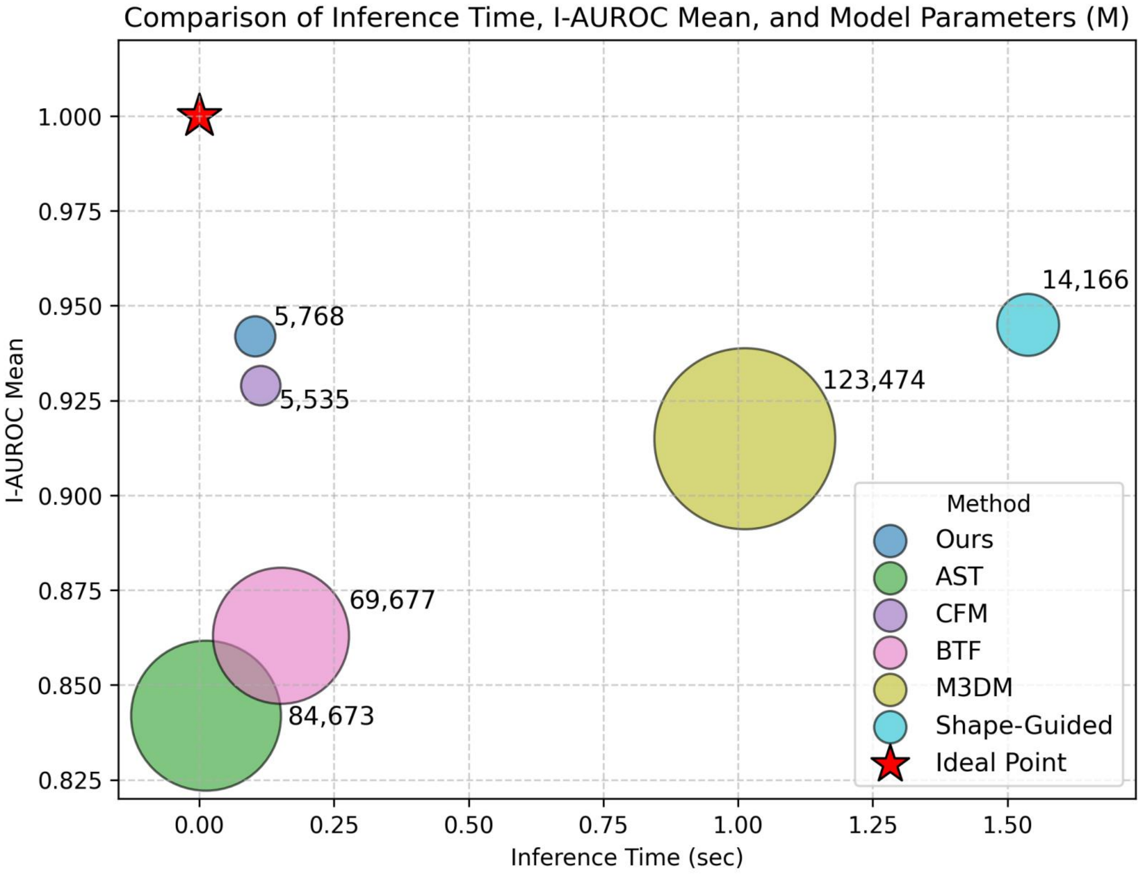
- The proposed method achieves the second-highest average AUC while maintaining a fast inference time, outperforming all other methods in speed-accuracy balance across multiple defect classes.

Model	Inference time	Mean	Bagle	Cable_gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire
Proposed method	**0.104**	**0.942 ± 0.004**	0.989 ± 0.005	0.853 ± 0.021	0.974 ± 0.007	0.990 ± 0.003	0.953 ± 0.014	0.881 ± 0.017	0.967 ± 0.007	0.924 ± 0.013	0.968 ± 0.011	0.922 ± 0.016
AST (2023, WACV) [6]	**0.012**	0.842 ± 0.003	0.959 ± 0.005	0.821 ± 0.029	0.842 ± 0.010	0.912 ± 0.011	0.966 ± 0.010	0.820 ± 0.014	0.796 ± 0.014	0.665 ± 0.012	0.990 ± 0.005	0.653 ± 0.015
CFM(2024, CVPR) [3]	0.114	0.929 ± 0.006	0.990 ± 0.004	0.849 ± 0.022	0.973 ± 0.005	0.992 ± 0.002	0.944 ± 0.006	0.846 ± 0.030	0.939 ± 0.025	0.907 ± 0.013	0.952 ± 0.004	0.897 ± 0.022
BTF (2023, CVPR) [4]	0.151	0.863 ± 0.007	0.911 ± 0.014	0.766 ± 0.036	0.960 ± 0.007	0.898 ± 0.024	0.923 ± 0.014	0.622 ± 0.031	0.849 ± 0.022	0.922 ± 0.013	0.920 ± 0.006	0.856 ± 0.025
M3DM (2023, CVPR) (Fail) [5]	1.012	0.915 ± 0.015	0.988 ± 0.011	0.829 ± 0.029	0.969 ± 0.013	0.959 ± 0.019	0.902 ± 0.042	0.899 ± 0.032	0.921 ± 0.024	0.929 ± 0.024	0.939 ± 0.017	0.815 ± 0.049
Shape-Guided Dual-Memory Learning (2023, ICML) (Fail) [2]	1.538	**0.945 ± 0.006**	0.990 ± 0.006	0.872 ± 0.017	0.974 ± 0.004	0.992 ± 0.002	0.978 ± 0.005	0.854 ± 0.013	0.985 ± 0.004	0.967 ± 0.005	0.965 ± 0.005	0.874 ± 0.017

Experimental Results

Performance Comparison (Trade-off Between Accuracy and Inference Time)

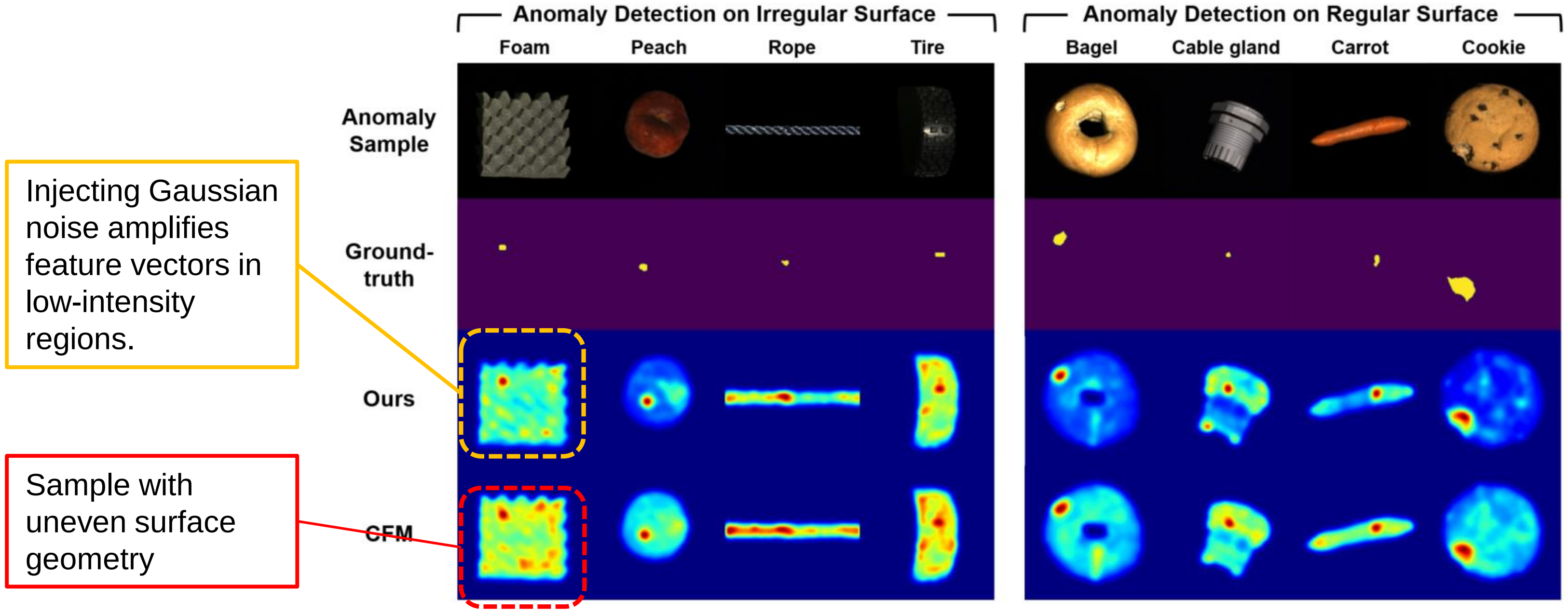
- The proposed model achieves the best trade-off, offering high accuracy with low inference time and minimal model size, closely approaching the ideal point in the design space.



Experimental Results

In-depth Analysis

- The proposed method effectively enhances anomaly signals in low-intensity regions using Gaussian noise, leading to precise localization even on geometrically irregular surfaces.
- Compared to CFM, the proposed model better suppresses noise and consistently detects fine-grained defects on both regular and uneven surfaces.



< Figure 13. > Anomaly Detection Results on Irregular and Regular Surfaces

Conclusion and Contribution

- The study primarily focuses on significantly optimizing inference speed and enhancing detection accuracy by employing sophisticated, advanced techniques with high-resolution 3D point cloud data in industrial settings.
- The proposed method addresses the challenge of detecting anomalies on irregular industrial surfaces by introducing a learnable Gaussian noise injection strategy that selectively enhances weak geometric features in low-intensity regions.
- Compared to conventional methods, the proposed approach demonstrates superior anomaly localization performance while achieving competitive inference time, as validated by statistical test results across multiple object categories.

Achievement

- Conferences
  - Unsupervised Learning-Based Anomaly Detection in Industrial Images Using 3D Point Clouds: A Survey, IEIE, 2024

Thank you

Reference

[1] Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2020). Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 14–19, 2020, pp. 4183–4192.

[2] Chu, Y.-M., Liu, C., Hsieh, T.-I., Chen, H.-T., & Liu, T.-L. (2023). Shape-guided dual-memory learning for 3D anomaly detection. In Proceedings of the 40th International Conference on Machine Learning, Honolulu, HI, USA, July 23–29, 2023, pp. 6185–6194.

[3] Costanzino, A., Ramirez, P. Z., Lisanti, G., & Di Stefano, L. (2024). Multimodal industrial anomaly detection by crossmodal feature mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 17–21, 2024, pp. 17234–17243.

[4] Horwitz, E., & Hoshen, Y. (2023). Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, June 18–22, 2023, pp. 12345–12354.

[5] Wang, Y., Peng, J., Zhang, J., Yi, R., Wang, Y., & Wang, C. (2023). Multimodal industrial anomaly detection via hybrid fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, June 18–22, 2023, pp. 23456–23465.

[6] Rudolph, M., Wandt, B., & Rosenhahn, B. (2023). Asymmetric student-teacher networks for industrial anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, January 3–8, 2023, pp. 34567–34576.