MulSen-AD: A Dataset and Benchmark for Multi-Sensor Anomaly Detection

Wenqiao Li[1*] **Bozhong Zheng**[1*] **Jinye Gan**[1] **Fading Lu**[1] **Xiang Li**[1] **Xiaohao Xu**[2]
Na Ni[1] **Zheng Tian**[1] **Shenghua Gao**[1†] **Yingna Wu**[1†]
[1]ShanghaiTech University [2]University of Michigan, Ann Arbor
{zhengbzh2023, liwq2022}@shanghaitech.edu.cn

Abstract

Object anomaly detection is an important problem in industrial quality inspection. Although remarkable progress has been made in object-level anomaly detection for data captured by a single type of sensor, two critical problems still hinder the deployment of the algorithms. First, in practical industrial environments, anomalies are often long-tailed and complex, making it inherently difficult for a single type of sensor to detect certain types of defects. For instance, color defects are invisible in 3D laser scanner, while subtle texture defects are also difficult to detect by infrared thermography sensor. Second, environment conditions, such as ambient lighting in factories, often greatly affect the detection process, leading to unreliable or invalid results under a single type of sensor. To address these issues, we propose the first Multi-Sensor Anomaly Detection (Mulsen-AD) dataset and develop a comprehensive benchmark (Mulsen-AD Bench) on Mulsen-AD dataset. Specifically, we build Mulsen-AD dataset with high-resolution industrial camera, high-precision laser scanner and lock-in infrared thermography. To exactly replicate real industrial scenarios, Mulsen-AD dataset is comprised of 15 types of distinct and authentic industrial products, featuring a variety of defects that require detection through the integration of multiple sensors. Additionally, we propose Mulsen-TripleAD algorithm, a decision level gating method, to solve the problem of object anomaly detection with the combination of three sensors. In our benchmark, our method achieved 81.1% object-level AUROC detection accuracy using three sensors, significantly outperforming single-sensor methods, demonstrating the potential value of multi-sensor datasets for the research community. The data and code(`https://github.com/ZZZBBBZZZ/MulSen-AD`) are released openly to support the development of the multi-sensor anomaly detection field.

1 Introduction

Object anomaly detection [8, 17] identifies defective objects by leveraging a pre-trained model to compare test objects with normal ones encountered during training. Most of prior works only considered detection setting under a single sensor [3, 4, 19, 16, 40], overlooking the potential benefits of utilizing data from multiple sensors for anomaly detection. However, in actual factory conditions, both the industrial environments and the anomalies are often complex. **Therefore, Object**

*Equal contribution.
†Corresponding author.

Submitted to the 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks. Do not distribute.

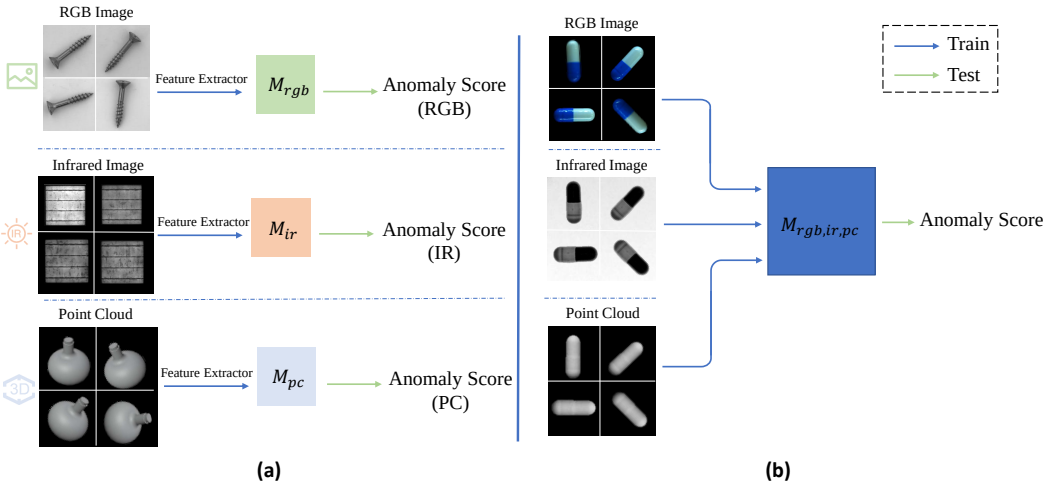


Figure 1: (a) Existing single-sensor object-level anomaly detection; (b) Our introduced multi-sensor object-level anomaly detection setting. M_{rgb}, M_{ir}, M_{pc} represent the pre-trained AD models for RGB, Infrared (IR), and Point Cloud (PC), and $M_{rgb,ir,pc}$ refers to a multi-sensor fusion AD model.

anomaly detection with a single type of sensor can fail to detect all anomalies due to the inherent limitations of single sensor. In fact, in many modern industrial factories, attempts are already being made to conduct object anomaly detection using multiple sensors. There is a solid need to propose a high-quality multi-sensor anomaly detection dataset to tap the gap between the academy and industry. In this work, we build a multi-sensor anomaly detection dataset (Mulsen-AD) and conduct a comprehensive benchmark on our Mulsen-AD dataset. Specifically, our Mulsen-AD dataset consists of 15 types of real industrial products, covering over 2000 objects, each containing RGB images, infrared images, and point cloud data. Furthermore, our dataset is also configured with diverse and realistic defects, which should be detectabel at least by one of the sensor data in our dataset. Additionally, we have developed a decision level gating method named Mulsen-TripleAD that meets with the need of multi-sensor anomaly detection task. To highlight the challenge of our AD setting and expedite research efforts towards developing more effective and universal multi-sensor anomaly detection approaches, we constructed a comprehensive benchmark with the combination of multi-sensor data, referred to Mulsen-AD Bench.

Overall, the main contributions of this paper are:

- We introduced Multi-Sensor Anomaly Detection (Mulsen-AD) setting, a practical and challenging setting for object level anomaly detection based on three different industrial sensors, solving the limitations of single sensor anomaly detection, and bridged the gap between industrial and research in object anomaly detection.

- We developed Mulsen-AD dataset, the first real dataset for evaluating multi-sensor anomaly detection, featuring diverse, high-quality object data captured by high-resolution devices and anomalies based on actual factory conditions.

- We conducted a comprehensive benchmark using the Mulsen-AD dataset and provided a universal toolkit to facilitate further exploration and reproduction of the benchmark.

- We proposed Mulsen-TripleAD, a decision fusion gating method for multi-sensor anomaly detection. Utilizing data from three types of sensors and combining multiple memory banks with a decision gating unit, Mulsen-TripleAD outperforms setups using fewer sensors and sets a new baseline for our multi-sensor anomaly detection task.

Table 1: Comparison between the proposed MulSen-AD and existing object-level anomaly detection datasets. Syn, IR, D, and PC represent Synthesis, Infrared, Depth, and Point Cloud.

Dataset	Year	Type	Modality	Sample Statistics			
				#Class	#Anomaly Types	#Sample	Multi-Sensor
MVTec-AD [4]	2019	Real	RGB	15	73	5354	✗
BTAD [19]	2021	Real	RGB	3	3	2830	✗
MPDD [13]	2021	Real	RGB	6	8	1346	✗
VisA [40]	2021	Real	RGB	12	-	10821	✗
MVTec LOCO-AD [3]	2022	Real	RGB	5	89	3644	✗
MAD [39]	2023	Syn+Real	RGB	20	3	4902	✗
Real-IAD [30]	2024	Real	RGB	30	8	151050	✗
GDXray [18]	2015	Real	X-ray	5	15	19407	✗
PVEL-AD [27]	2023	Real	IR	1	10	36543	✗
MVTec3D-AD [5]	2021	Real	RGB-D	10	3-5	3604	✗
Eyecandies [6]	2022	Syn	RGB-D	10	3	15500	✗
Real3D-AD [16]	2023	Real	PC	12	2	1200	✗
Anomaly-ShapeNet [14]	2023	Syn	PC	40	6	1600	✗
MulSen-AD (Ours)	2024	Real	RGB/IR/PC	15	14	2035	✓

2 Related work

Object-level Anomaly Detection Datasets. Object-level anomaly detection aims to identify defective samples during or after industrial production processes. Historically, this field has relied solely on data from single sensor. MVTec-AD [4], BTAD [19], MPDD [13], and VisA [40] is a series of single view photo-realistic industrial anomaly detection datasets. The objects provided in these datasets are just captured in one single view RGB camera. The overall shape information of objects cannot be captured, and texture information is easily affected by lighting and environmental conditions in this setting. The LOCO AD dataset [3] provides rich global structural and logical information but is not suitable for fine-grained anomaly detection on individual objects. MVTec3D-AD [5] and Eyecandies [6] intend to integrate depth maps with RGB maps to provide the geometry information under the fixing single view. MAD [39] and Real-IAD [30] are multi-view AD datasets, trying to provide texture information and depth information from different views. Visual anomaly detection under RGB cameras cannot avoid being easily influenced by ambient lighting and the confusion in detecting superficial morphological abnormalities. As a solution for this, datasets such as PVEL-AD [27] and GDXray [18] intend to detect external and shallow layer anomalies within an object by infrared and X-ray sensors. Nonetheless, PVEL-AD and GDXray sacrifice color and texture information. Real3D-AD [16] and Anomaly-ShapeNet [14] are 3D AD datasets, which only focus on object-level geometry anomaly detection. In a word, existing object AD datasets just rely on one single kind of sensor, which often fail to accurately capture all types of anomalies in actual factory settings, significantly limiting the advancement of this area. To address this issue and explore the setting of multi-sensor anomaly detection AD problem, we propose the first multi-sensor anomaly detection(MulSen-AD) dataset, which includes RGB images, infrared images, and high-resolution point clouds. Table 1 lists the comparison between Mulsen-AD and other object AD datasets.

Multi-sensor Datasets. Many multi-sensor datasets are available in the field of autonomous driving. It is important to note that early autonomous driving was also constrained by the limitations of single sensor. Initial autonomous driving datasets primarily focused on annotating images captured by RGB cameras only, such as Cityscapes [10], Synthia [24], Vistas [20], and BDD100K [36]. However, achieving human-level autonomous driving requires precise perception and localization in the real 3D world, and relying solely on 2D images for this purpose is inadvisable [33]. To enable robust perception in the real 3D world, multi-sensor datasets have been developed. These datasets typically include not only camera images but also spatial information from sensors like radar, LiDAR, and even thermal information from infrared cameras. The KITTI dataset [11] is a comprehensive dataset in autonomous driving research, providing a variety of data types including images from stereo cameras, LiDAR point clouds, GPS positioning, and IMU data. Subsequently, more recent multi-sensor datasets such as NuScenes [7], Waymo [28], and V2V4Real [33] have emerged, offering a larger number of scenes and further expanding the depth and diversity of data available for autonomous

driving research. Multi-sensor datasets have significantly enhanced the accuracy and reliability of autonomous driving. Similarly, this technology holds promise for achieving more precise detection and localization of anomalies in the industrial field. By integrating visual, thermal, and spatial information from multi-sensor, there is potential to improve anomaly detection in industrial products.

Multi-sensor Fusion Methods. Existing multi-sensor fusion methods can be categorized into data (early) fusion, feature (middle) fusion, and decision (late) fusion [1]. Data fusion methods, such as PointPainting [29], PointAugmenting [31], MVP [35], and RVF-Net [21] aim to integrate data from various sources or sensors early, creating a unified representation that can be directly utilized for subsequent processing steps. Feature fusion methods such as DeepFusion [15], TransFusion [2], EPNet [12], AutoAlignV2 [9] and DeepInteration [34] strive to facilitate the transformation of input data into more abstract feature representations across different layers in the training phase, empowering the model to effectively utilize data at each network layers. Late fusion methods, including CLOCs [22] and RangeLVDet [38], follow a strategy where data from multi-sensor is processed independently before being combined at the fusion stage, aiming to minimize errors caused by discrepancies in the data.

3 MulSen-AD Dataset Construction

3.1 Sensor selection

The most common sensors in industrial scenarios include industrial cameras, infrared sensors, ultrasonic sensors and 3D laser scanners. The understanding of signal of ultrasonic sensors strongly relies on the physical knowledge, since there are several different noises which should be eliminated. Therefore, we did not involve ultrasonic signal in our datasets. MulSen-AD includes RGB images acquired by cameras, gray-scale images by lock-in infrared thermography and high-resolution 3D point clouds by laser scanners. Three kinds of sensor are explained in the following.

1) RGB camera. The Daheng MER2-230-168U3C camera, with a maximum resolution of 1920x1200, was mounted on a UR5 robotic arm to capture images above the target object. Line light sources were positioned on either side of the object to ensure uniform lighting, minimizing shadows and enhancing the visibility of intricate surface details. This light setup also allowed us to simulate complex lighting conditions in a real industrial environment and capture multi-light images.

2) Lock-in infrared thermography. Noverlteq TWILIS-180 lock-in infrared system with FLIR A600 infrared camera (640x480 resolution) was applied to capture the IR gray-scale images. To detect anomalies, periodic thermal stimulation is applied to heat the objects. If the heat absorption of anomalies is different from the objects, the temperature difference is presented in the images. In our dataset, infrared camera successfully captures the temperature anomaly of broken inside capsule, damages in solar panels, detachment of parts inside lamps and the other internal anomalies that RGB camera cannot detect.

3) 3D scanner sensory. Creaform MetraSCAN 750, the scanning system with HandyPROBE portable CMM, is applied to obtain high-precise point cloud data. This system consists of two parts, which are hand-held MetraSCAN-3D scanner and dual-camera C-Track sensor. During the scanning process, MetraSCAN-3D projects multiple laser crossing on the objects and captures the reflected light to present 3D surface of the object. The global coordinates are generated by C-Track through continuously monitoring the positions of MetraSCAN-3D and scanned object. With the precision of 0.03 mm and resolution of 0.05 mm, the system captures 20k to 100k points per object. The hand-held design of system allows operators to scan objects from all angles, which eliminates the blind areas that may exist captured by the fixed-position scanning devices, such as the Zivid One-Plus (applied by MVTec3D-AD) and PMAX-S130 (applied by Real3D-AD). With this portable system, we detected 3D geometric anomalies, including deformations of spring pad and creases in cotton. It is challenging to detect these anomalies by RGB and infrared cameras. Detailed specifications of these sensors and descriptions of the corresponding anomalies are included in Table 2 and Figure 2.

Table 2: Data collection equipments parameters.

Device	Daheng MER2-230-168U3C	Noverlteq TWILIS-180	Creaform MetraSCAN 750
Modality	RGB image	Infrared image	Point cloud
Resolution	1920×1200	640×480	0.05 mm
Accuracy	—	$\pm 2°C$	0.03 mm
Pixel Depth	8bit	16bit	—
Wavelength Range	—	$7.5–14\,\mu m$	—
Scanning Area	—	—	$275 \times 250\,mm$

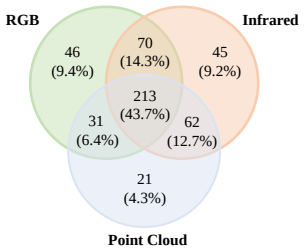


Figure 2: Distribution of anomaly collected by individual and multiple sensors in MulSen-AD. Overlap areas indicate anomalies detected by multiple sensors.

3.2 Material Preparation

We selected 15 objects made by different materials, including metal, plastic, fiber, rubber, semiconductor and composite materials, with different shapes, sizes and colors. Additionally, we manually created 14 types of anomalies, including cracks, holes, squeeze, external and internal broken, creases, scratches, foreign bodies, label, bent, color, open, substandard, and internal detachments, as shown in Figure 3. The anomalies are designed to closely resemble real industrial situations, with a wide distribution of types, including surface, internal, and 3D geometric anomalies.

3.3 Data Collection and Process

Firstly, the infrared camera was positioned on the top of centrally placed objects, which was oriented at random horizontal angles. Then, thermal stimulation sources periodically heated the object for 30 to 180 seconds, depending on the difference of material and thickness. Infrared images were captured at a resolution of 640×480 in gray scale. Subsequently, the RGB camera with the resolution of 1280x960 recorded images from the top view. Based on the previously captured infrared images, the position of objects and height of camera were adjusted through measuring scale, UR5 robotic arm and the grid of camera software. Finally, the point clouds were acquired by laser scanner. The objects were flipped during scanning process to ensure the integrity of 3D point clouds, which is called "dual-scan strategy". The point clouds from each scanning were first coarsely aligned manually, followed by fine alignment using the Iterative Closest Point (ICP) algorithm. This process typically needs to be repeated multiple times until the point cloud is accurately aligned. The collection pipeline is illustrated in Figure 3. Collection details are provided in the supplementary Sec 1.

3.4 Data Annotation

After data collection, we manually mark pixel-level masks of anomalies in RGB and infrared images via LabelMe. For point clouds, we use Geomagic Design X to select the anomalous regions and then save the anomalous points in txt format. It is noteworthy that there will be no corresponding masks for that modality, if the anomaly of an object is not visible in RGB images, infrared images or point clouds. For example, if a capsule with an internal anomaly is detected in the infrared image, while undetected in the RGB image or point clouds, only the infrared image is marked.

3.5 Data Statistics

Table 3 presents the statistical information of MulSen-AD dataset, which includes the dataset category, the number of training set samples, the number of normal and abnormal samples in the test set, and the number of anomaly types. The MulSen-AD dataset comprises 2035 samples evenly distributed across 15 categories. Figure 4-(b) displays the types of anomalies associated with each category, with an average of 4.8 defect types per category. As detailed in Figure 4-(a), it is obvious that

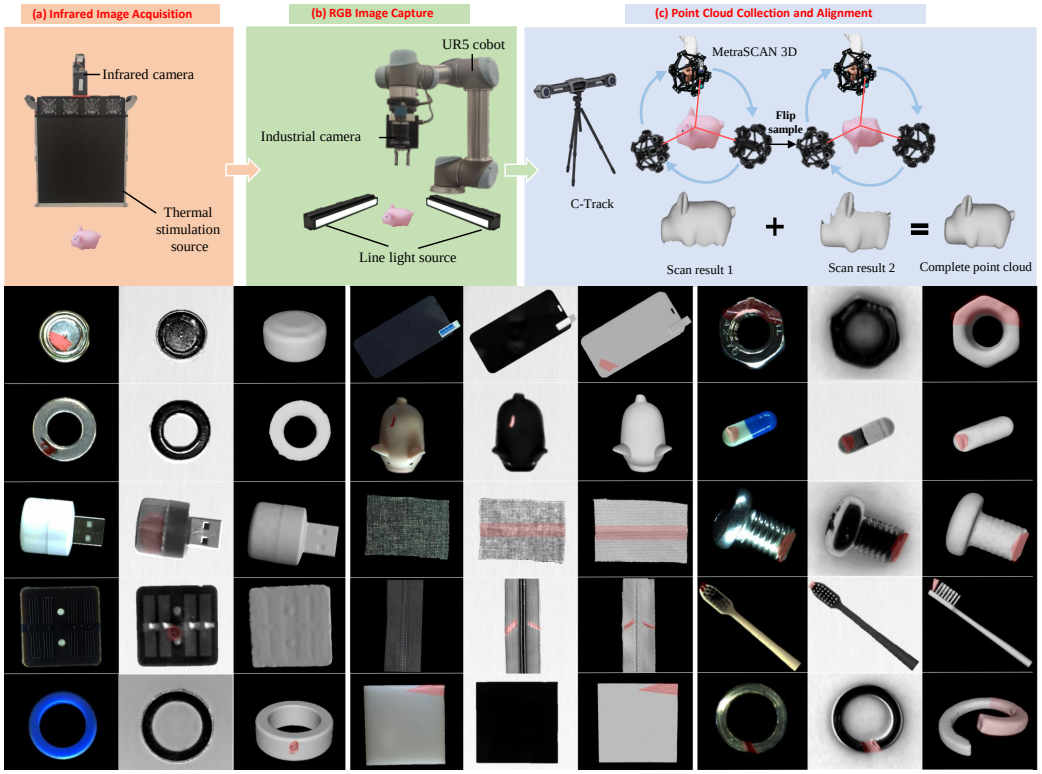


Figure 3: Data collection pipeline for the proposed MulSen-AD dataset consists of three steps: (a) Infrared image capture, (b) RGB image capture, and (c) Point cloud collection and alignment. The bottom of the figure shows partial visualizations of 15 categories from the MulSen-AD dataset, displaying three modalities (RGB, IR, PC) for each category. Parts of the detects are only visible in single or two modalities. Abnormal areas are highlighted in red.

the three modalities show different advantages in detecting anomalies in different categories. For instance, anomalies on solar panels are most effectively identified in infrared images, while anomalies on cotton are more accurately detected using point clouds. Venn diagram, as shown in Figure 2, visually indicates the intersection and differentiation in capability of defecting detection for different sensors. In the non-overlapping regions, each sensor demonstrates its unique ability to identify specific anomalies. In the overlapping regions, the detection capability of different sensors is comparable. It indicates that combining different sensors' information can enhance the anomaly detection accuracy. The demonstrated diversity and complementarity of data by multiple sensors in Mulsen-AD underscore the importance of implementing multi-sensor fusion algorithms to improve object anomaly detection capabilities.

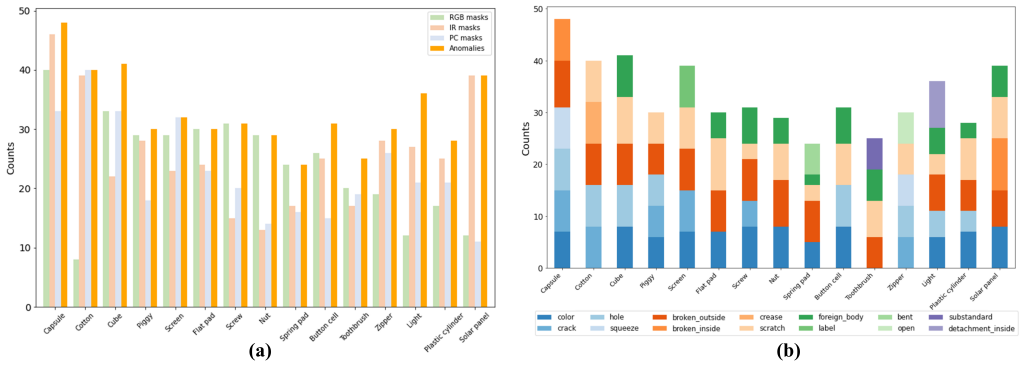
4 Mulsen-AD Benchmark and Baseline

4.1 Problem Definition of Multi-Sensor Anomaly Detection

Our Multi-Sensor Anomaly Detection (Mulsen-AD) setting is introduced for object-level anomaly detection task, which can be stated as follow: Given a training set $\mathcal{T} = \{t_i\}_{i=1}^{N}$, in which $\{t_1, t_2, \cdots, t_N\}$ are the anomaly-free samples and each sample t consists of a RGB image I_{rgb}, an infrared image I_{ir} and a 3D point cloud P. In addition, $\mathcal{T}$ belongs to one certain category c_j, $c_j \in \mathcal{C}$, where $\mathcal{C}$ denotes the set of all categories. During test time, each test sample contains data $\{I_{\text{rgb}}, I_{\text{ir}}, P\}$ from three sensor and their corresponding modality labels $\{L_{\text{rgb}}, L_{\text{ir}}, L_{\text{P}}\}$. If more than one of the modality label in $\{L_{\text{rgb}}, L_{\text{ir}}, L_{\text{P}}\}$ equals to 1, the object will be regarded as anomalous object and be assigned an object level label L_{o} equalling to 1. Otherwise, the object label will be

Table 3: Statistics of MulSen-AD dataset. Note that the proportions of abnormal pixels and points are calculated only in abnormal samples.

Category	Train Set	Test Set		#Total	#Anomaly	Abnormal Pixel/Point Ratio[%]		
	#Normal	*#Normal*	*#Abnormal*		Classes	RGB	Infrared	Point Cloud
Capsule	64	10	48	122	6	0.392	0.346	11.1
Cotton	78	10	40	128	5	0.771	0.569	3.32
Cube	110	10	41	161	5	0.558	0.552	2.07
Piggy	110	10	30	150	5	0.444	0.444	1.37
Screen	69	10	32	111	4	0.774	1.07	4.28
Flat pad	90	10	30	130	4	0.188	0.193	4.99
Screw	90	10	31	131	5	0.314	0.330	3.48
Nut	118	10	29	157	4	0.201	0.117	5.85
Spring pad	86	10	24	120	5	0.056	0.078	19.1
Button cell	90	10	31	131	4	0.259	0.227	1.69
Toothbrush	110	10	25	145	5	0.105	0.126	7.58
Zipper	86	10	30	126	5	0.687	0.997	6.19
Light	110	10	36	156	6	0.209	0.838	1.24
Plastic cylinder	90	10	28	128	5	0.317	0.427	1.94
Solar panel	90	10	39	139	5	0.306	0.458	0.496
Mean	93	10	33	136	4.8	0.372	0.451	4.98
Total	1391	150	494	2035	72	—	—	—



(a) (b)

Figure 4: Distribution statistics of anomalies and masks in the MulSen-AD dataset. The number of masks in each modality also represents the number of anomaly samples visible within each modality. (a) Distribution of total anomalies and mask counts per modality for each category. (b) Distribution of data volume across different defect categories.

0, which means this object is a normal sample. Given a test object (normal or abnormal) and its object label L_o from category c_j, the Mulsen-AD model should discriminate whether the object is anomalous.

4.2 Mulsen-TripleAD

Inspired by Patchcore [25] and M3DM [32], we propose a universally applicable multi-sensor anomaly detection method named Mulsen-TripleAD, which is illustrated in Figure 5. The Mulsen-TripleAD consists of three important parts: (1) Multiple Feature extraction: two pretrained feature extractors, DINO [37] for RGB&Infrared Images and PointMAE [23] for point clouds are used to extract different representations respectively. (2) Memory Bank establishment: we denote three sensor memory banks as M_{rgb}, M_{ir}, M_{pc} respectively. We refer Patchcore [25] to build these memory banks with normal samples during training time. During inference, each sensor memory bank is used to predict an sensor anomaly score. (3) Decision Gating Unit: after obtaining each sensor image level anomaly score, we draw on the learnable One-Class Support Vector Machine (OCSVM) [26] from the M3DM [32] as our decision gating unit $\mathcal{D}_a$ to integrate anomaly scores from different sensor and obtain the final object-level anomaly score a, which can be described as:

$$a = \mathcal{D}_a(\phi(\mathcal{M}_{rgb}, f_{rgb}), \phi(\mathcal{M}_{pt}, f_{pt}), \phi(\mathcal{M}_{ir}, f_{ir})), \tag{1}$$

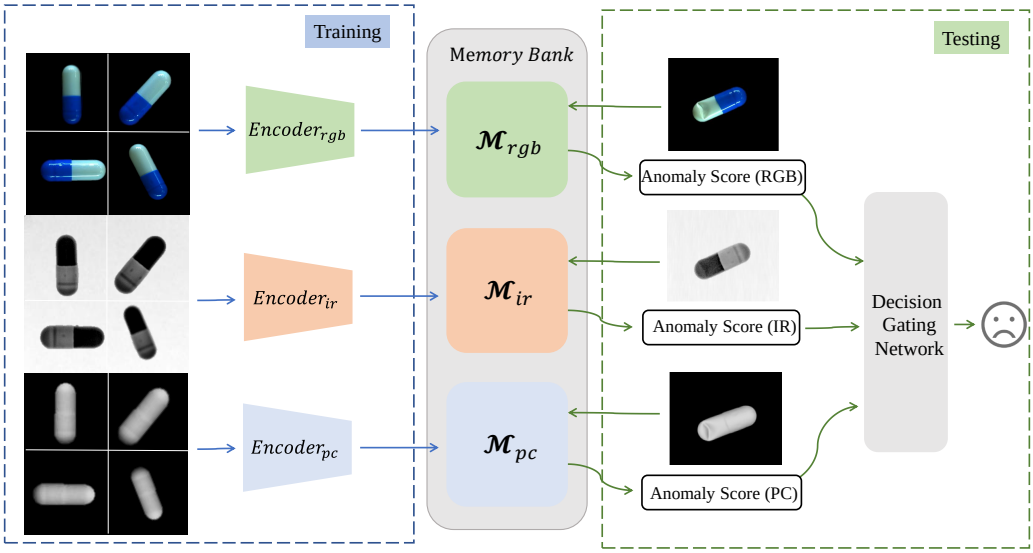


Figure 5: The pipeline of Mulsen-AD. Our Mulsen-AD contains three important parts:(1) Multiple Feature extraction;(2) Memory Bank establishment;(3) Decision Gating Network. $\mathcal{M}_{rgb}$, $\mathcal{M}_{ir}$, and $\mathcal{M}_{pc}$ represent memory banks for RGB, Infrared, and Point Cloud.

where ϕ are the score functions introduced by [25], which can be formulated as:

$$\phi(\mathcal{M}, f) = \eta \| f^{(i,j),*} - m^* \|_2, \tag{2}$$

$$f^{(i,j),*}, m^* = \arg \max_{f^{(i,j)} \in f} \arg \min_{m \in \mathcal{M}} \| f^{(i,j)} - m \|_2, \tag{3}$$

where $\mathcal{M} \in \{\mathcal{M}_{rgb}, \mathcal{M}_{pt}, \mathcal{M}_{ir}\}$, $f \in \{f_{rgb}, f_{pt}, f_{ir}\}$, f represent the features extracted from rgb images, infrared images or point clouds during inference and η is a re-weight parameter.
More Mulsen-TripleAD implementation details are listed in our supplementary Sec 5.

4.3 Mulsen-AD Bench

Benchmarking Method Selection. To comprehensively investigate the performance of different sensor data fusion in our multi-sensor anomaly detection setting, we adopt our Mulsen-TripleAD algorithm in different kinds of combination of sensor as our benchmark methods. The corresponding benchmark experiments are presented in Table 4. In Table 4, Single refer to that we just adopt the RGB or Infrared images or point cloud into our pipeline without decision gating. Double refers to that we combine two of our three sensor data into our Mulsen-TripleAD algorithm pipeline. Triple means the original version, which we have illustrated in Figure 5. To obtain great representations, we selected pretrained PointMAE [23] and DINO [37] as feature extractors for point cloud, RGB and infrared data. The memory bank setting refers to Patchcore [25] and the decision gating unit adopt the setting of M3DM [32]. It's important to note that we set the parameters all the same between our experiments to ensure the fair comparison.

Metric. Following previous work, We specifically choose the Area Under the Receiver Operating Characteristic Curve (AUROC) as the metric for evaluating the performance of anomaly detection at the object-level.

Code and Toolkit. To complement Mulsen-AD Bench for more people, we release a comprehensive toolkit including code and dataset for multi-sensor anomaly detection, which includes all the single, double and triple core methods we have mentioned above.

Table 4: MULSENAD BENCH on Mulsen AD dataset. The score indicates object-level AUROC ↑. PC means Point cloud. The best result of each category is highlighted in bold.

Category	Single			Double			Triple
	RGB	*Infrared*	*PC*	*RGB+Infrared*	*PC+RGB*	*PC+Infrared*	*PC+RGB+Infrared*
Capsule	0.927	0.754	0.625	0.923	0.942	0.775	**0.958**
Cotton	0.720	0.818	0.795	0.830	0.742	**0.835**	0.762
Cube	0.947	0.960	0.338	**0.978**	0.947	0.949	**0.978**
Spring pad	**0.900**	0.542	0.446	0.575	0.875	0.538	0.712
Screw	0.148	**0.910**	0.465	0.794	0.152	0.890	0.639
Screen	0.412	0.972	0.612	0.912	0.444	**0.984**	0.872
Piggy	0.980	0.903	0.537	0.983	0.980	0.903	**0.997**
Nut	**0.814**	0.103	0.483	0.138	0.790	0.103	0.252
Flat pad	**0.813**	0.417	0.747	0.543	**0.813**	0.400	0.610
Plastic cylinder	0.843	**1.000**	0.579	**1.000**	0.821	**1.000**	**1.000**
Zipper	0.637	0.982	0.595	0.946	0.699	**1.000**	0.943
Button cell	**1.000**	0.797	0.468	0.923	**1.000**	0.790	0.971
Tooth brush	0.587	0.852	0.875	0.852	0.670	**0.898**	0.822
Solar panel	0.528	**0.923**	0.374	0.821	0.482	0.877	0.836
Light	0.814	**0.839**	0.244	0.814	0.789	**0.839**	0.808
Average	0.738	0.776	0.546	0.802	0.743	0.785	**0.811**

Analysis of Mulsen-AD Bench. From Table 4, it is clearly shown that our baseline method, Mulsen-TripleAD, outperforms other anomaly detection methods for the Mulsen-AD dataset. Specifically, our MulSen-Triple3D method achieves an accuracy of 81.1% by combining three sensors, which surpasses the average single sensor result of 68.7% (Average of the single sensor results for RGB, Infrared and PC) and the average double sensors result of 77.7% (Average of the double sensor results for RGB+Infrared, PC+RGB and PC+Infrared), which demonstrates that the reasonable integration of data from multiple sensors enhances anomaly detection performance under the MulSen-AD setting. In addition, in Table 4, significantly different anomaly detection results for certain category under different single sensor further validates the appropriateness of our Mulsen-AD dataset and setting.

Single 3D Benchmark. Given the scarcity of real 3D anomaly detection dataset and the high quality of our Mulsen-AD data, we also tested the single 3D anomaly detection benchmarks on the 3D subset of Mulsen-AD dataset. The evaluation metrics include Image-AUROC and Pixel-AUROC, as in previous 3D anomaly detection tasks. Due to the space limit, the 3D AD results are listed in our supplementary Sec 3.

5 Discussion

Conclusion. In this work, we introduce Multi-Sensor anomaly detection setting, and develop the Mulsen-AD dataset, which is the first exploration to evaluate Multi-Sensor anomaly detection algorithms. In addition, we propose Mulsen-TripleAD to solve the problem of unsupervised object level anomaly detection with the combination of three sensors, which serves a baseline for our dataset.

Limitation&Future work. Although our MulSen-AD dataset provides RGB, infrared images, and point cloud data from three common industrial anomaly detection sensors, which are sufficient for most industrial scenarios, sensors like X-ray and ultrasound are also widely used in modern factories. Further integrating more types of high-quality sensor data to address complex industrial detection scenarios is the next stage where we need to continue our efforts. Additionally, we have currently only explored multi-sensor fusion methods at the decision level. Further exploration is needed on modality and feature-level fusion approaches.

Potential negative social impacts. Our dataset was collected with permission from the factory, so no negative social impact will exist.

References

[1] Simegnew Yihunie Alaba, Ali C Gurbuz, and John E Ball. Emerging trends in autonomous vehicle perception: Multimodal fusion for 3d object detection. *World Electric Vehicle Journal*, 15(1):20, 2024.

[2] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1090–1099, 2022.

[3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision*, 130(4):947–969, 2022.

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9584–9592, 2019.

[5] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2022.

[6] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the Asian Conference on Computer Vision*, pages 3586–3602, 2022.

[7] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[8] Yunkang Cao, Xiaohao Xu, Jiangning Zhang, Yuqi Cheng, Xiaonan Huang, Guansong Pang, and Weiming Shen. A survey on visual anomaly detection: Challenge, approach, and prospect. *arXiv preprint arXiv:2401.16402*, 2024.

[9] Zehui Chen, Zhenyu Li, Shiquan Zhang, Liangji Fang, Qinhong Jiang, and Feng Zhao. Deformable feature aggregation for dynamic multi-modal 3d object detection. In *European conference on computer vision*, pages 628–644. Springer, 2022.

[10] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

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

[12] Tengteng Huang, Zhe Liu, Xiwu Chen, and Xiang Bai. Epnet: Enhancing point features with image semantics for 3d object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, pages 35–52. Springer, 2020.

[13] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 66–71, 2021.

[14] Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. *arXiv preprint arXiv:2311.14897*, 2023.

[15] Yingwei Li, Adams Wei Yu, Tianjian Meng, Ben Caine, Jiquan Ngiam, Daiyi Peng, Junyang Shen, Yifeng Lu, Denny Zhou, Quoc V Le, et al. Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17182–17191, 2022.

[16] Jiaqi Liu, Guoyang Xie, Ruitao Chen, Xinpeng Li, Jinbao Wang, Yong Liu, Chengjie Wang, and Feng Zheng. Real3d-ad: A dataset of point cloud anomaly detection. *Advances in Neural Information Processing Systems*, 36, 2024.

[17] Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1):104–135, 2024.

[18] Domingo Mery, Vladimir Riffo, Uwe Zscherpel, German Mondragón, Iván Lillo, Irene Zuccar, Hans Lobel, and Miguel Carrasco. Gdxray: The database of x-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 34(4):42, 2015.

[19] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*. IEEE, June 2021.

[20] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulò, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5000–5009, 2017.

[21] Felix Nobis, Ehsan Shafiei, Phillip Karle, Johannes Betz, and Markus Lienkamp. Radar voxel fusion for 3d object detection. *Applied Sciences*, 11(12):5598, 2021.

[22] Su Pang, Daniel Morris, and Hayder Radha. Clocs: Camera-lidar object candidates fusion for 3d object detection. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10386–10393. IEEE, 2020.

[23] Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pages 604–621. Springer, 2022.

[24] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016.

[25] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022.

[26] Bernhard Schölkopf, Robert C Williamson, Alex Smola, John Shawe-Taylor, and John Platt. Support vector method for novelty detection. *Advances in neural information processing systems*, 12, 1999.

[27] Binyi Su, Zhong Zhou, and Haiyong Chen. Pvel-ad: A large-scale open-world dataset for photovoltaic cell anomaly detection. *IEEE Transactions on Industrial Informatics*, 19(1):404–413, 2023.

[28] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[29] Sourabh Vora, Alex H. Lang, Bassam Helou, and Oscar Beijbom. Pointpainting: Sequential fusion for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[30] Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jianning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection, 2024.

[31] Chunwei Wang, Chao Ma, Ming Zhu, and Xiaokang Yang. Pointaugmenting: Cross-modal augmentation for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11794–11803, June 2021.

[32] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8032–8041, 2023.

[33] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, et al. V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13712–13722, 2023.

[34] Zeyu Yang, Jiaqi Chen, Zhenwei Miao, Wei Li, Xiatian Zhu, and Li Zhang. Deepinteraction: 3d object detection via modality interaction. *Advances in Neural Information Processing Systems*, 35:1992–2005, 2022.

[35] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Multimodal virtual point 3d detection. *Advances in Neural Information Processing Systems*, 34:16494–16507, 2021.

[36] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.

[37] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

[38] Zehan Zhang, Zhidong Liang, Ming Zhang, Xian Zhao, Hao Li, Ming Yang, Wenming Tan, and Shiliang Pu. Rangelvdet: Boosting 3d object detection in lidar with range image and rgb image. *IEEE Sensors Journal*, 22(2):1391–1403, 2021.

[39] Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, and Hao Zhao. Pad: A dataset and benchmark for pose-agnostic anomaly detection, 2023.

[40] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation, 2022.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes]
- Did you include the license to the code and datasets? [No]
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We discussed it at Section 5.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] The dataset and code can be downloaded form `https://github.com/ZZZBBBZZZ/MulSen-AD`
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] All experiments were conducted on Tesla v100.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We cited the models used at Section 4.2
   (b) Did you mention the license of the assets? [Yes] All assets used is open-licensed.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Our baseline method and MulSen-AD datasets can be downloaded from `https://github.com/ZZZBBBZZZ/MulSen-AD`
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A] Our research is not related to human subjects

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]