Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection

Chengjie Wang[1,2]   Wenbing Zhu[3,4]   Bin-Bin Gao[2]   Zhenye Gan[2]

Jiangning Zhang[2]   Zhihao Gu[1]   Shuguang Qian[4]   Mingang Chen[5]   Lizhuang Ma[1*]

[1]Shanghai Jiao Tong University   [2]Youtu Lab, Tencent   [3]Fudan University   [4]Rongcheer Co., Ltd

[5]Shanghai Development Center of Computer Software Technology

{jasoncjwang, danylgao, wingzygan, vtzhang}@tencent.com, ma-lz@cs.sjtu.edu.cn

{Louis.zhu, Bruce.qian}@rongcheer.com, cmg@sscenter.sh.cn, ellery-holmes@sjtu.edu.cn

Website: https://realiad4ad.github.io/Real-IAD

Abstract

Industrial anomaly detection (IAD) has garnered significant attention and experienced rapid development. However, the recent development of IAD approach has encountered certain difficulties due to dataset limitations. On the one hand, most of the state-of-the-art methods have achieved saturation (over 99% in AUROC) on mainstream datasets such as MVTec, and the differences of methods cannot be well distinguished, leading to a significant gap between public datasets and actual application scenarios. On the other hand, the research on various new practical anomaly detection settings is limited by the scale of the dataset, posing a risk of overfitting in evaluation results. Therefore, we propose a large-scale, Real-world, and multi-view Industrial Anomaly Detection dataset, named Real-IAD, which contains 150K high-resolution images of 30 different objects, an order of magnitude larger than existing datasets. It has a larger range of defect area and ratio proportions, making it more challenging than previous datasets. To make the dataset closer to real application scenarios, we adopted a multi-view shooting method and proposed sample-level evaluation metrics. In addition, beyond the general unsupervised anomaly detection setting, we propose a new setting for Fully Unsupervised Industrial Anomaly Detection (FUIAD) based on the observation that the yield rate in industrial production is usually greater than 60%, which has more practical application value. Finally, we report the results of popular IAD methods on the Real-IAD dataset, providing a highly challenging benchmark to promote the development of the IAD field.

1. Introduction

High-quality datasets play a pivotal role in the development of computer vision technology, guiding the direction of technological advancement and bridging the gap between technical research and practical application. For instance, ImageNet [13] has made an indelible contribution to the development of deep learning model structures and learning methodologies. MVTec AD [2] abstracts defect detection in manufacturing into a research topic, thereby bringing visual learning algorithms closer to industrial production application scenarios. Nevertheless, the data volume of current AD datasets is still small and needs further development.

Industrial production is the cornerstone of human societal development, with product quality inspection playing a key role [7]. Defects in the production process of parts can affect product quality and subsequently reduce product lifespan. In the production of pharmaceuticals, food, batteries, and other products, product defects can pose a threat to human safety. Given the importance of defect detection, manufacturing companies have invested substantial resources in this area [2–4], and academia has also gradually turned its attention to this issue in recent years [1, 12, 15, 25, 26, 34]. Some automation technologies have begun to play a role in practical applications. Early applications of visual learning technology in industrial defect detection mainly involved supervised learning for detection [6, 14, 28] and segmentation [8, 16, 32] tasks. Although these methods have practical value, they still face technical challenges such as the need for precise defect location annotation. Moreover, the detection performance significantly decreases for defects that are scarce or not included in the training set. With the advent of datasets like MVTec AD [2] and VisA [40], a large number of unsupervised anomaly detection methods have emerged [12, 23, 26, 29, 34, 38]. These techniques only need to ensure that the images in the

---

*Corresponding authors.

training set are anomaly-free, and resulting models have the ability to predict defect locations and pixel areas, significantly reducing manual annotation costs and endowing the algorithm with the ability to recognize unknown defects.

The emergence of datasets like MVTec AD [2], and VisA [40] has stimulated academic interest in industrial anomaly detection research and has given birth to many innovative methods. However, with technological progress, these datasets have gradually revealed their limitations. For example, the defect range in MVTec AD is small, and the actual application scenarios are simple. Recent methods have exceeded 99% in I-AUROC (image-level) and P-AUROC (pixel-level) metrics, making it difficult to distinguish the merits of new methods. The scale of the data is also limited, with some categories in MVTec AD having only about 60 or even fewer defect images, and the random error caused by the small number of test images cannot be ignored. Some new IAD research uses some defective images in the training set, resulting in a smaller test set and exacerbating the problem. In terms of object diversity, MVTec AD includes 15 types of objects, and VisA includes 12 types of objects. The limited number of object types affects the evaluation of the capabilities of a unified model, leading to overly optimistic model metrics. In addition, there is a gap between the current dataset setup and the ideal application scenario. Although IAD is unsupervised learning, the training set still requires manual annotation. This process may introduce noise samples, such as the noise samples in the BTAD training set proposed in SoftPatch [21]. Furthermore, mainstream 2D IAD datasets consist of single-view images, but in practical applications, part structures are complex, and a single view cannot cover all defects. Although some datasets, like MVTec AD-3D [4], attempt to solve the multi-view problem from a 3D perspective, the high cost of 3D sensors limits them in practical applications.

To make the dataset closer to actual application scenarios and address the limitations of existing datasets, we propose a new **Real**-world **I**ndustrial **A**nomaly **D**etection dataset called Real-IAD. This dataset far exceeds existing datasets in terms of the number of object categories and the number of images. To our knowledge, this dataset is the first to consider the multi-view problem in 2D IAD tasks. In an ideal scenario for practical application, image acquisition equipment is deployed on the production line, and the algorithm is trained automatically without human intervention, possessing the ability to determine whether a product is defective. Based on the characteristic that the yield rate of most production lines is greater than 60%, we add 0∼40% of quality inspection anomaly images to the anomaly-free train set, proposing the completely unsupervised anomaly detection problem for the first time. Through this new dataset, we hope to promote technological development in the field

of industrial anomaly detection, encourage the emergence of more efficient and practical detection methods, and provide stronger technical support for industrial production. In this new real-world dataset, we have collected 30 industrial products covering a variety of materials, such as plastic, wood, ceramics, and mixed materials *etc.*, considering the diversity in industrial production. We have also deliberately increased the resolution of the dataset to capture more subtle defect features, providing strong support for high-precision defect detection. In addition, we have collected images of each product from multiple angles to address the problem that a single view cannot cover all defects. To verify the effectiveness of the new dataset and promote technological development, we conducted a series of experiments on the dataset. Specifically, we conduct benchmark tests on existing unsupervised anomaly detection algorithms to assess their performance on the new dataset. The experimental results show that although these algorithms have achieved good performance on the original datasets, there is still room for improvement on the new dataset, indicating that the new dataset has a certain level of challenge, which can help promote algorithm improvement and innovation.

In summary, our main contributions are as follows:

- We propose a new Real-IAD dataset, which is more than ten times larger than existing mainstream datasets. It includes 30 classes of objects with each containing 5 shooting angles, totaling 150K high-resolution images. Moreover, Real-IAD presents more challenging defects with a larger range of defect area and ratio proportions, better differing the performance between different methods and meeting various research settings of IAD.
- We construct a Fully Unsupervised IAD setting on the Real-IAD dataset that is closer to actual application scenarios, where only naturally existing constraint is used that the yield rate of most production lines is greater than 60%, without introducing additional manual annotations.
- We report the performance of popular IAD methods on the Real-IAD dataset in several settings and provide a highly challenging benchmark to promote the development of the anomaly detection field.

2. Related Work

2.1. Anomaly Detection Datasets

Early anomaly detection works are generally conducted on KolektorSDD [31] which only contains one category and greatly limits the evaluation and development of algorithms. Subsequently, datasets such as MTD [19], MPDD [20] and BTAD [27] are proposed, but these datasets are still relatively small, in terms of the number of categories and total number of images. Since the introduction of the MVTec AD dataset [2], the conventional sensory IAD tasks have started to get on track and gradually attracted a large num-

ber of researchers and practitioners. This dataset contains 15 industrial products in 2 types with a total of 5,354 images, greatly promoting algorithm research. Later VisA [40] covers 12 objects in 3 types with a total of 10,821 images, elevating the number of IAD dataset to the 10K level for the first time while having a larger number of categories, *i.e.*, 15. Recently, Zhou *et al*. [39] propose a synthetic PAD to promote research on pose-agnostic anomaly detection, but there is a natural gap between synthetic data and real samples, leading to inconsistencies in the metrics. In addition, some datasets have extended 3D information to better detect defects, *e.g*., MVTec 3D AD [4], Eyecandies [5], and Real3D [24]. Nevertheless, the current visual IAD datasets are still small in scale and category, generally applied to limited industrial scenarios, and there is no IAD dataset equivalent to the status of ImageNet-1K [13] in the classification field. To address this issue, we propose a novel large-scale (approximately 150K), more categories (30 categories), and multi-view (5 shooting angles) Real-IAD dataset for the task, which contain high-definition images, elaborate labelings, and more challenging detects, committed to providing a more challenging benchmark to promote the development of IAD methods and fair comparison of different methods (*c.f*., Sec. 4 for details).

2.2. Standard Anomaly Detection

The standard IAD task aims to identify whether a given image of a target class is anomalous or not and localize the anomaly region if the prediction is anomalous. The lack of anomalous data makes it more challenging and it is usually formulated as an unsupervised learning problem with only normal data available in the training phase. Recently, vast efforts are dedicated to developing unsupervised anomaly detectors and form several mainstream approaches, namely data augmentation-based methods [22, 36], reconstruction-based methods [17, 23], and embedding-based methods [12, 26, 29, 30]. Specially, the embedding-based methods can be further divided into four categories, *i.e*., memory bank [29], normalizing flow [30], knowledge distillation [12, 37], and classification [26]. These methods produce superior results.

2.3. Other Settings in Anomaly Detection

More challenging settings have been proposed one after another. The zero-/few-shot IAD [9, 10, 18] focuses on employing a few normal samples for IAD, reducing the demand for data. SoftPatch [21] introduces noisy data (less than 10%) into the anomaly-free training data in the standard setting to model the real scenario, and Zaheer *et al*. propose unsupervised video detection [35]. Contrarily, the semi-supervised IAD [33] introduces anomalous data during training for better IAD. To avoid re-training for different categories, the unified IAD [34] accomplished IAD for multiple classes by a unified framework. In this paper, we

propose a new setting called Fully Unsupervised Industrial Anomaly Detection (FUIAD) based on the observation that the yield rate in industrial production is usually greater than 60%, which has a more practical application value.

3. Real-IAD Dataset Description

3.1. Data Collection and Construction Manner

This section introduces the construction pipeline of Real-IAD in Fig. 1, and the procedure includes three parts:

1) **Material Preparation.** We have purchased and collected 30 objects covering a variety of materials, including metal, plastic, wood, ceramics, and mixed materials. Some objects are shown in the upper part of Fig. 1-(a). We manually created various types of defects, including missing parts, dirt, deformation, pits, damage, holes, cracks, scratches, as shown in the lower part of Fig. 1-(a). Subsequently, these normal and anomaly objects are sent to the prototype machine for image collection.

2) **Prototype Construction.** The construction of the acquisition equipment is shown in Fig. 1-(b). There are a total of five cameras capturing the object from different angles, with one camera taking top shots and the other four taking shots from symmetric angles at approximately 45 degrees. Additionally, to better image the object and clearly capture minor defects, a ring light source is installed above the object. In actual applications, automated equipment is used to flip the parts for bottom quality inspection. For parts with more structural obstructions, the camera is rotated to capture images from more angles. Considering that part flipping and additional shooting angles may affect actual applications but not algorithm research, we abstract the multi-view visual quality inspection problem in real scenarios into five shooting angles. The camera adopted is HIKROBOT MV-CE200-10GC with $3{,}648 \times 5{,}472$ resolution.

3) **Data Collection, Annotation, and Cleaning.** As shown inFig. 1-(c), to ensure the accuracy of dataset collection, we first manually confirm all normal images and anomaly images and use LabelMe to manually label pixel-level masks for anomaly data. Then, the data is divided into three groups, each of which is supervised trained using a cascade RCNN based on HRNet-32w backbone for confident learning. We look for inconsistencies between the model's prediction results and the manual annotation results. These data will undergo manual inspection and annotation again until the model's predicted AP results remain essentially unchanged, and the number of annotated images that need to be modified is less than a certain threshold. Finally, we consider the data to be clean and construct different dataset settings.
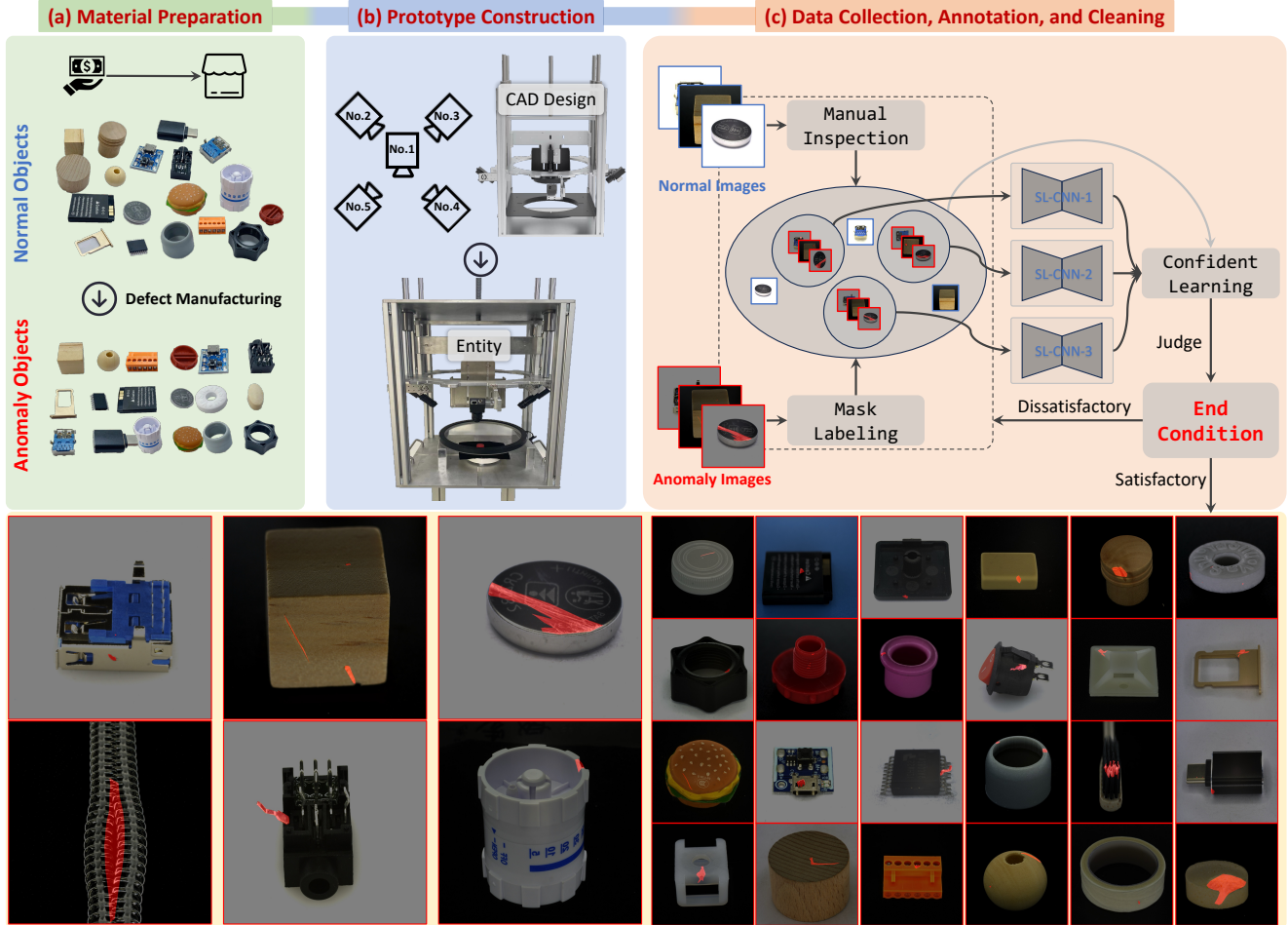
Figure 1. **Data collection pipeline for our proposed Real-IAD dataset,** which consists of four steps in tandem: **(a)** Material preparation and defect manufacturing. **(b)** Prototype design and construction that contains 5 shots for capturing multi-view images simultaneously, *i.e.*, one top-down camera with extra four cameras arranged uniformly at 45 degrees. **(c)** Collection procedure that includes cyclical processes of data collection, annotation, and cleaning. At the bottom, partial visualization of the final Real-IAD dataset reveals that the proposed Real-IAD exhibits large scale (30 classes), a wide range of defect proportions (0.01% to 6.75%), and a broad defect ratio (1:1 to 1:10), indicating that Real-IAD is highly challenging. Abnormal areas are prominently marked in red.

Table 1. **Comparison with current popular 2D real-world IAD datasets on different attributes. ✔: Satisfied. ✗: Unsatisfied.**

Dataset	Class	Image Number			Image Resolution	Defect Types Per Category	Segmentation Labeling	Multiple Shooting Angles
		Normal	Anomaly	All				
BTAD	3	952	392	1344	600∼1,600	3.00	✔	✗
SSGD	1	0	2504	2504	1,500×1,000	7.00	✗	✗
MVTec AD	15	4,096	1,258	5,354	700∼1,024	4.87	✔	✗
VisA	12	9,621	1,200	10,821	960∼1,562	7.50	✔	✗
Read-IAD	30	99,721	51,329	151,050	2,000∼5,000	4.37	✔	✔

3.2. Comparison with Popular 2D Datasets

This section provides a detailed statistical analysis of the proposed Real-IAD dataset, and compares it with mainstream datasets, *i.e.*, MVTec AD [2] and VisA [40].

Comparison with Mainstream Datasets. As shown in Tab. 1, compared to mainstream datasets, our method has at least a 2x increase in the number of categories, and a magnitude increase in quantity, *i.e.*, from 10K to 150K. Furthermore, Real-IAD provides images with a resolution of up to 2,000∼5,000 to support research on finer-grained IAD algorithms, which involves cropping from original images.
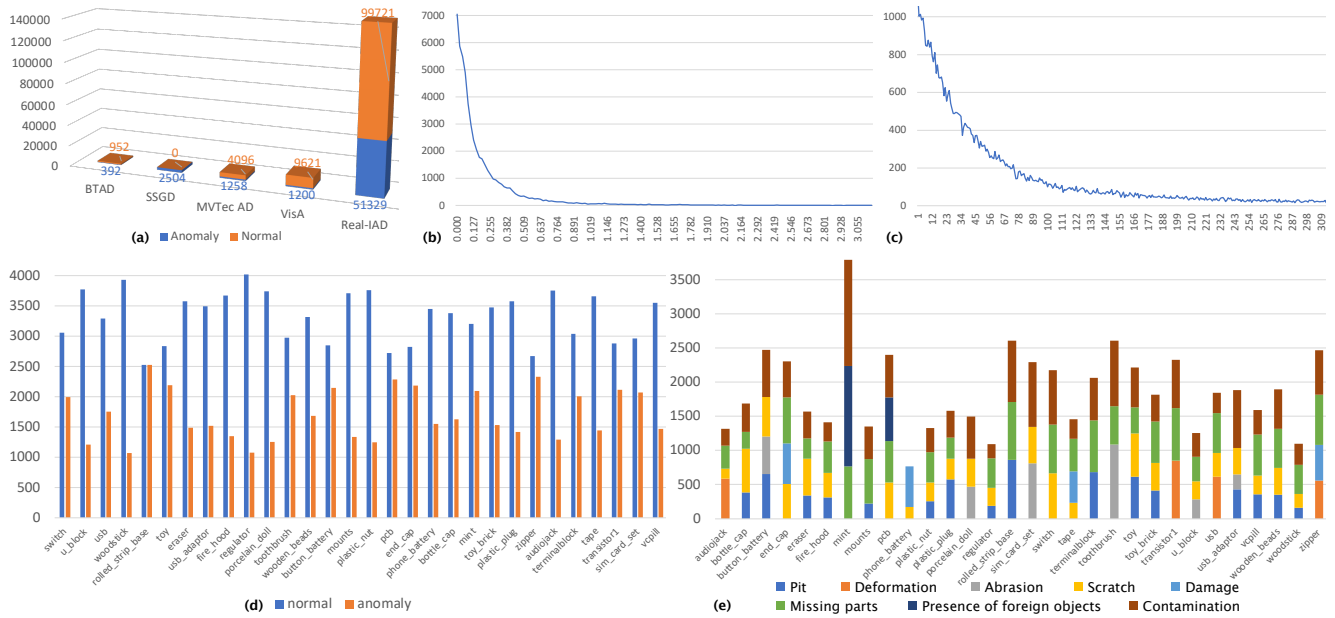
Figure 2. **Statistic information of our proposed Real-IAD datset.** *(a)* Statistic comparison of anomaly/normal data volume among popular datasets. *(b)* Statistics of the percentage of the image area occupied by the anomaly region. *(c)* Statistics of the aspect ratio of the minimum bounding rectangle of the defect. *(d)* Distribution of anomaly/normal image quantities across different categories. *(e)* Distribution of data volume across different defect categories.
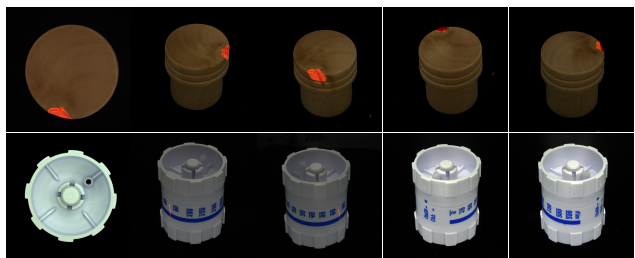

Figure 3. **Visualization of multi-view collections**, where the first column represents the results captured by the top-down camera, and the other columns represent the results captured by four surrounding cameras at 45 degrees. Abnormal images are highlighted for visualization, and normal images are displayed with normal brightness. It can be seen that multi-view shooting can solve the problem of invisibility of abnormal areas from a single viewpoint.

Furthermore, each object is provided with five images taken from different angles with segmentation labeling to support a multi-view setting, *c.f.*, Fig. 3.

Data Statistics. Fig. 2 shows statistic of our Real-IAD dataset. Compared to the contrastive datasets, our Real-IAD has an order of magnitude improvement on both normal and abnormal data (Fig. 2-a). At the same time, the proportion of defect area (Fig. 2-b) and the range of defect ratio (Fig. 2-c) are larger, indicating a higher difficulty level of the dataset, which is also proven by the experiments in Tab. 2. The bottom of Fig. 2 displays the number of normal and abnormal data in different categories (Fig. 2-d), as

well as the proportion of different types of defects (Fig. 2-e). Overall, the ratio of normal to abnormal data is not very disparate, *e.g.*, VisA reaches 8:1, and each type of category has multiple types of defects.

Advantage Analyses. Our Real-IAD dataset has advantages in the following aspects: *1)* **Diversity**: Compared with existing datasets, the Real-IAD dataset covers a wider range of categories and provides richer scenarios, which helps to train more robust anomaly detection models and conduct fair evaluations. *2)* **Large-scale**: Real-IAD is the first to provide a dataset of over 150K images, a magnitude increase compared to popular IAD datasets, which also provides multi-view images with pixel-level annotations. *3)* **Challenge**: Compared with existing datasets, the Real-IAD has a higher level of difficulty, which can drive the development and progress of current anomaly detection algorithms.

3.3. Real-IAD Visualization

The bottom of Fig. 1 displays 30 sample data types from Real-IAD, demonstrating that the dataset includes data from a variety of material types, such as metal, plastic, wood, ceramics, and mixed material. Moreover, the types of defects include pit, deformation, abrasion, scratches, damage, missing parts, foreign objects, and contamination. The defect area occupies a proportion ranging from 0.1 to 0.5, with a ratio ranging from 0.1 to 10.0. The richer data indicates that Real-IAD is highly challenging, and it is expected to promote further development in IAD.

4. Benchmark

4.1. Evaluation settings

Unsupervised IAD (UIAD): The algorithm learns patterns and structures only from normal samples. Therefore, most existing anomaly datasets are usually divided into training and testing sets, where the training set only contains normal samples. In contrast, the testing set consists of both normal and anomalous samples. Following the previous arts, we fisrt evaluate our Real-IAD with the UIAD setting.

Fully Unsupervised IAD (FUIAD): The setting is inevitable in real-world IAD but is seldom discussed. And it is difficult to construct the FUIAD setting due to the limited availability of anomalous samples in the existing datasets, such as MVTec AD [2] and VisA [40]. This paper is the first to consider FUIAD (*i.e.*, UIAD with noise) from a dataset construction perspective. In the Real-IAD, large and diverse samples of anomalies give us the flexibility to build an FUIAD setup.

To create an FUIAD setting, we first need to fix the testing set including normal and anomaly samples. In our experiment, the number of normal and anomaly samples is uniformly set to 100 samples (500 images). The remaining normal and anomaly samples will be a candidate set for constructing a noisy training set. We keep the total number of samples in the training set constant but the corresponding numbers of normal and abnormal ones are adaptively adjusted with a given noise ratio. To do this, we need first infer the number of training samples according to the scale of the candidate set and the range of noisy ratio (e.g., [0.1, 0.4]). Then, we randomly sample normal and anomaly samples from the candidate set to construct a training set with a specific noisy ratio. In this way, we obtain several new fully unsupervised benchmarks with different noise ratios dubbed $\alpha \in [0, 1]$.

4.2. Evaluation Metric

The AUROC is the most widely used metric for image-level and pixel-level anomaly detection. In addition, a normalized Per-Region Overlap (PRO) between segmentation and ground truth is calculated and the Area Under PRO curve (AUPRO) is also adopted as a pixel-level metric [2].

Most IAD methods [12, 22, 26, 29, 30, 36, 38] evaluate anomaly detection only at image and pixel levels due to limitations of existing IAD datasets. The proposed Real-IAD, to the best of our knowledge, is the first multi-view anomaly detection dataset, where each sample consists of multiple different views. Therefore, in addition to evaluating anomaly detection performance at the image and pixel levels, we integrate the results from multiple views to assess sample-level performance. This is more consistent with the indicator evaluation in industrial production lines.

5. Comparisons with IAD Benchmarks

5.1. Results on Unsupervised IAD

We comprehensively compare the performance of our Real-IAD with some popular IAD benchmarks (e.g., MVTec AD [2] and VisA [40]). MVTec AD [2] is a widely used dataset in the field of industrial anomaly detection. It consists of high-resolution images of 10 objects and 5 textures, captured under different lighting conditions and with different types of anomalies. The dataset provides ground truth annotations for the location and type of anomalies present in the images, enabling quantitative evaluation of detection performance. VisA [40] is $2\times$ larger than MVTec AD, with both image and pixel-level annotations. It spans 12 objects across 3 domains, with challenging scenarios including complex structures in objects, multiple instances, and object pose/location variations. The proposed Real-IAD is a large-scale (approximately 150K) with more categories (30 objects), and multi-view (5 shooting angles) dataset for anomaly detection, where each anomaly image is labeled by a pixel-level mask and a specific defect type. Considering that MVTec and VisA are single-view, we also select one viewpoint (*i.e.*, top view) from Real-IAD to form a single-view Real-IAD, a subset of multi-view Real-IAD.

For the UIAD setting, we mainly choose embedded-based IAD methods, such as PatchCore[29], PaDim [11], and CFlow [30], data-augmentation-based methods, SimpleNet [26] and DeSTSeg [38], and reconstruction-based RD [12] and UniAD [34] for performance comparisons. PatchCore [29] first extracts neighborhood-aware patch-level features and then stores them in a memory bank. At test time, images are classified as anomalies if at least one patch is anomalous, and pixel-level anomaly segmentation is generated by scoring each patch feature. PaDiM [11] extract pre-trained features to model normal distribution, then utilize a distance metric to measure the anomalies. CFlow [30] proposed to use a conditional normalizing flow framework to estimate the exact data likelihoods which are infeasible in other generative models for IAD. SimpleNet [26] and DeSTSeg [38] convert Unsupervised IAD into supervised IAD training with generated anomaly images/features and real normal images/features. Feature reconstruction-based RD [12] and UniAD [34] prevent the model from learning the shortcut with reverse distillation and neighborhood-masked attention, respectively. We reproduce PatchCore [29], PaDim [11], CFlow [30] and RD [12] with open-sourced Anomalib, and other methods including SimpleNet [26], DeSTSeg [38] and UniAD [34] with official codes. In our experiments, we resize all images to 256×256, only center crop 224×224 from the resized 256×256 for PatchCore [29] and PaDim [11]. Other hyperparameters, such as batch size and learning rate, are kept the same as the official implementation.

Table 2. UIAD performance comparisons on Real-IAD and some existing IAD datasets, MVTec AD [2] and VisA [40]. The results are averaged over all categories on each dataset. We report the averaged value and standard deviation (denoted as "mean±std") based on the results of different methods for each metric. The lower the value means the corresponding anomaly detection task is more challenging.

Datasets	Metric↑	Embedding-based			Data-Aug-based		Reconstruction-based		Mean±Std
		PatchCore[29]	PaDim[11]	CFlow[30]	SimpleNet[26]	DeSTSeg[38]	RD[12]	UniAD[34]	
MVTec AD [2]	I-AUROC	99.0	95.5	97.1	99.3	98.4	98.5	97.5	97.9 ±1.2
	P-AUPRO	94.7	91.3	92.3	89.9	94.7	93.1	90.7	92.4 ±1.8
VisA [40]	I-AUROC	95.1	85.1	91.8	95.4	90.3	95.9	88.8	91.8 ±3.7
	P-AUPRO	91.2	79.7	84.8	88.7	90.6	92.9	85.5	87.6 ±4.2
Real-IAD (single-view)	I-AUROC	90.4	86.6	85.0	91.7	89.2	87.6	82.7	87.6 ±2.9
	P-AUPRO	92.5	90.0	88.6	88.9	90.3	94.4	86.0	90.0 ±2.6
Real-IAD (multi-view)	S-AUROC	93.7	91.2	89.8	94.9	94.0	83.7	88.1	90.3 ±4.0
	I-AUROC	89.4	80.3	82.5	88.5	86.9	87.1	82.9	85.0 ±3.6
	P-AUPRO	91.5	81.8	90.5	84.6	81.5	93.8	86.1	86.3 ±4.4

The results of all methods on MVTec, VisA, and our single-/multi-view Real-IAD are presented in Table 2. 1) We can observe that there is a significant performance decrease from MVTec (97.9% in I-AUROC) to our single-view and multi-view Real-IAD (85% in I-AUROC). This indicates that the proposed Real-IAD is more challenging than the existing datasets for anomaly detection. The performance drop is more pronounced when using a unified model, which is reasonable considering the more complex data distribution (*i.e.*, more views and more classes) in Real-IAD; 2) It is hard to evaluate different methods on the existing datasets as the results are very similar. Especially on MVTec, most methods achieve about 98%-99% in I-AUROC. In contrast, on the Real-IAD dataset, most methods only obtain about 90% I-AUROC, which is better for evaluating the effectiveness of anomaly detection algorithms; 3) The pixel-level P-PRO on Real-IAD is comparable to the existing VisA but noticeably lower than MVTec AD. This indicates that the proposed Real-IAD dataset also presents challenges at the pixel-level anomaly localization.

5.2. IAD Benchmarks Meet FUIAD

Traditional Unsupervised IAD methods indeed assume that the training dataset only contains normal samples. However, in practical applications, it isn't easy to guarantee that all training samples are normal. Considering that the yield rate of good products on a production line is typically higher than 60%, it means that no more than 40% products are abnormal. Therefore, existing unsupervised methods may not be suitable for practical applications because they require manual annotation before training that is laborious.

Hence, a more practical solution is to perform fully unsupervised anomaly detection, which allows a certain proportion of abnormal samples in the training process. This paradigm can better adapt to real-world scenarios and reduce reliance on manual annotation. By building a fully Unsupervised IAD setting, it is possible to better simulate abnormal situations in real-world scenarios, thereby improving the robustness and practicality of IAD algorithms.

Table 3. Comparison with popular IAD datasets on FUIAD setting. NR (α) and # TS are the noisy ratio and anomalous testing samples, respectively. The fail settings are marked in blue.

NR	# TS	# valid category			# training images or samples			
		MVTec AD	ViSA	Real-IAD	MVTec AD	ViSA	Real-IAD(I)	Real-IAD(S)
$\alpha = 0.1$	25	15	12	30	268	667	2605	521
	50	13	12	30	223	500	2466	493
	75	9	12	30	200	250	2327	465
	100	4	0	30	181	0	2188	437
	150	0	0	30	0	0	1910	382
$\alpha = 0.2$	25	15	12	30	252	375	2931	586
	50	13	12	30	193	250	2774	554
	75	9	12	30	148	125	2618	523
	100	4	0	30	120	0	2462	492
	150	0	0	30	0	0	2149	429
$\alpha = 0.4$	25	15	12	30	152	187	3908	781
	50	13	12	30	108	125	3699	739
	75	9	12	30	78	62	3491	698
	100	4	0	30	60	0	3283	656
	150	0	0	30	0	0	2866	573

Popular anomaly detection datasets like MVTec AD [2] and VisA [40] are primarily designed for unsupervised settings. To evaluate the capabilities of fully unsupervised algorithms, researchers have to randomly sample a portion of anomaly samples from the testing set and add them as noise samples to the training set, creating a fully unsupervised IAD experiment. In this way, it leads to a significant reduction in the number of testing samples, which may be insufficient to effectively evaluate the performance of anomaly detection algorithms. For quantitative analysis, we assume a certain number of testing samples (e.g., 25, 50, 75, 100 and 150) for each category and calculate the number of valid testing categories and the number of training images or samples based on a given noise ratio (e.g., 0.1, 0.2 and 0.4). The corresponding statistic results are reported in Table 3.

We can see that increasing the noisy ratio or injecting anomalous noisy samples into the normal training set (MVTec and VisA) would lead to a decrease in the number of valid testing categories and the number of normal training samples, making it impossible to evaluate the performance of FUIAD algorithms in some challenging settings (high noisy ratios). In contrast, the proposed Real-IAD ensures a consistent number of valid testing categories and a stable scale of corresponding normal samples in the training set when the noisy ratio ranges from 0.1 to 0.4 and the num-

Table 4. FUIAD performance (S-AUROC/I-AUROC/P-PRO) comparisons on Real-IAD. The results are averaged over 30 categories. The best results are marked in red and the second-best ones are marked in blue.

Settings	Embedding-based				Data-Aug-based		Reconstruction-based	
	PaDim[11]	CFlow[15]	PatchCore[29]	SoftPatch[21]	SimpleNet[26]	DeSTSeg[38]	RD[12]	UniAD[34]
$\alpha = 0.0$	92.7 / 84.6 / 84.4	88.8 / 83.9 / 90.6	94.4 / 91.3 / 92.6	93.9 / 91.4 / 92.1	94.9 / 89.8 / 83.9	94.6 / 89.6 / 88.7	85.5 / 89.3 / 95.0	89.6 / 85.4 / 87.6
$\alpha = 0.1$	85.5 / 81.9 / 86.4	81.8 / 80.3 / 90.7	92.2 / 90.4 / 93.2	92.8 / 90.9 / 92.9	87.8 / 83.3 / 79.9	88.1 / 85.6 / 86.9	83.5 / 88.1 / 95.1	87.2 / 84.2 / 87.7
$\alpha = 0.2$	82.2 / 80.1 / 86.5	81.3 / 79.6 / 90.7	91.1 / 89.5 / 93.0	92.2 / 90.5 / 92.9	82.7 / 79.6 / 75.9	81.4 / 80.3 / 83.2	80.7 / 87.3 / 94.9	85.2 / 82.8 / 87.3
$\alpha = 0.4$	76.5 / 77.0 / 86.1	78.6 / 78.0 / 90.2	88.9 / 88.1 / 92.4	90.6 / 89.3 / 92.5	77.7 / 74.7 / 70.5	74.5 / 74.4 / 75.5	76.8 / 84.5 / 94.7	81.5 / 80.1 / 86.6

ber of anomalous testing samples varies from 50 to 150. We believe that Real-IAD provides a more reliable and comprehensive assessment for FUIAD algorithms.

6. Comparisons with Fully Unsupervised IAD

For fully unsupervised anomaly detection, we set noisy ratio $\alpha \in \{0.1, 0.2, 0.4\}$. In addition, to make a fair comparison between FUIAD and UIAD settings on the same testing set, we also set the noisy ratio to 0, which transforms FUIAD into UIAD. The results of FUIAD ($\alpha = 0$) should be the upper bound of all FUIAD settings ($\alpha > 0$). Considering that it is not easy to establish an effective fully unsupervised setting on existing datasets, we only evaluate fully unsupervised anomaly detection methods on our Real-IAD. Similar to the evaluation of unsupervised anomaly detection, we also select the same methods, PatchCore[29], PaDim [11], CFlow [30], SimpleNet [26], DeSTSeg [38], RD [12] and UniAD [34]. To alleviate the effect of noise samples, SoftPatch [21] proposes a denoising mechanism based on PatchCore for memory bank construction. Therefore, we also evaluate the SoftPatch method under the FUIAD setting on our Real-IAD. The main results are reported in Tabel 4.

Under the setting of unsupervised anomaly detection (the noisy ratio is set to zero), most state-of-the-art anomaly detection methods (such as PatchCore [29], SoftPatch [21], SimpleNet [26], DeSTSeg [38]) show almost similar performance for sample- and image-level anomaly classification. For pixel-level anomaly segmentation, PatchCore [29], SoftPatch [21] and RD [12] that use multi-level features are more advantageous, thanks to the low-level features retaining rich spatial location information.

In the setting of fully unsupervised anomaly detection, almost all methods suffer severe performance degradation on all metrics, especially PaDim [11], SimpleNet [26], DeSTSeg [38] and RD [12]. Unsupervised PatchCore [29] shows robustness, thanks to the patch-level Memory-Bank mechanism. In the Memory-Bank, normal and abnormal features are stored simultaneously, and stable detection can be achieved as long as the distribution of abnormal features during the inference phase is different from the stored anomalies. In addition, the proportion of abnormal pixels is actually very low even with a high noise ratio at sample-level (such as 0.4), and thus the quality of normal patch features in Memory-Bank can be ensured to some extent. The

SoftPatch [21] that first filters some noisy features and then constructs Memory-Bank achieves almost optimal performance in all methods. However, the performance improvement is relatively limited compared to unsupervised Patch-Core [29]. This means that fully unsupervised anomaly detection still requires more in-depth research efforts. *E.g.*, using model ensemble methods to improve robustness to noise, combining the semantic understanding capabilities of large vision language models to provide prior distribution of noisy data, *etc*.

7. Conclusion

Based on the analysis of existing anomaly detection datasets and practical industry applications, we have several observations. First of all, unsupervised industrial anomaly detection algorithms have almost reached saturation in performance, but it is still hard to deploy them in actual industrial inspection applications. Secondly, there is a gap between algorithms and applications, where most algorithms rely on clean normal training samples, but the data obtained in actual industrial production lines contains a certain amount of noise. Finally, fully unsupervised industrial anomaly detection is a more suitable setting for practical applications, but existing datasets are not sufficient to support research due to limited samples. In order to address the above problems, we propose a large-scale, real-world, multi-view anomaly detection dataset (Real-IAD), which contains 150k high-resolution images, covering 30 objects of metal, plastic, wood, ceramics, and mixed materials, each object contains 5 different views with 8 common defects. The Real-IAD provides high-quality annotations at different levels such as pixel, image, and sample. We established two settings, unsupervised and fully unsupervised, to conduct extensive and comprehensive evaluations using state-of-the-art anomaly detection methods. We hope that the Real-IAD can advance research in the field of anomaly detection.

Limitation and Future Works. This paper only reports the results of some typical methods on Real-IAD. In the future, we will replicate more methods on Real-IAD and provide results under more settings, *e.g.*, zero-shot, few-shot, and semi-supervised settings, *etc*. In addition, considering the large-scale and multi-view characteristics of Real-IAD, it is also worthwhile to further study algorithms that are well-suited to these features.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, pages 622–637. Springer, 2019. 1

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1, 2, 4, 6, 7

[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, 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, VISIGRAPP 2022, Volume 5: VISAPP, Online Streaming, February 6-8, 2022*, pages 202–213. SCITEPRESS, 2022. 1, 2, 3

[5] 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. 3

[6] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6154–6162, 2018. 1

[7] 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. 1

[8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 1

[9] Xuhai Chen, Yue Han, and Jiangning Zhang. A zero-/few-shot anomaly classification and segmentation method for cvpr 2023 vand workshop challenge tracks 1&2: 1st place on zero-shot ad and 4th place on few-shot ad. *arXiv preprint arXiv:2305.17382*, 2023. 3

[10] Xuhai Chen, Jiangning Zhang, Guanzhong Tian, Haoyang He, Wuhao Zhang, Yabiao Wang, Chengjie Wang, Yunsheng Wu, and Yong Liu. Clip-ad: A language-guided staged dual-path model for zero-shot anomaly detection. *arXiv preprint arXiv:2311.00453*, 2023. 3

[11] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 6, 7, 8

[12] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9737–9746, 2022. 1, 3, 6, 7, 8

[13] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1, 3

[14] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 1

[15] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 98–107, 2022. 1, 8

[16] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1

[17] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8791–8800, 2021. 3

[18] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *ECCV*, pages 303–319. Springer, 2022. 3

[19] Yibin Huang, Congying Qiu, and Kui Yuan. Surface defect saliency of magnetic tile. *The Visual Computer*, 36:85–96, 2020. 2

[20] 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. IEEE, 2021. 2

[21] Xi Jiang, Jianlin Liu, Jinbao Wang, Qiang Nie, Kai Wu, Yong Liu, Chengjie Wang, and Feng Zheng. Softpatch: Unsupervised anomaly detection with noisy data. *Advances in Neural Information Processing Systems*, 35:15433–15445, 2022. 2, 3, 8

[22] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9664–9674, 2021. 3, 6

[23] Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *IEEE Transactions on Image Processing*, 2023. 1, 3

[24] 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. *arXiv preprint arXiv:2309.13226*, 2023. 3

[25] Jiaqi Liu, Guoyang Xie, Jingbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep in-

dustrial image anomaly detection: A survey. *arXiv preprint arXiv:2301.11514*, 2, 2023. 1

[26] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20402–20411, 2023. 1, 3, 6, 7, 8

[27] 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)*, pages 01–06. IEEE, 2021. 2

[28] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 1

[29] 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. 1, 3, 6, 7, 8

[30] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1088–1097, 2022. 3, 6, 7, 8

[31] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3):759–776, 2020. 2

[32] Tete Xiao, Yingcheng Liu, Bolei Zhou, Yuning Jiang, and Jian Sun. Unified perceptual parsing for scene understanding. In *Proceedings of the European conference on computer vision (ECCV)*, pages 418–434, 2018. 1

[33] Minghui Yang, Peng Wu, and Hui Feng. Memseg: A semi-supervised method for image surface defect detection using differences and commonalities. *Engineering Applications of Artificial Intelligence*, 119:105835, 2023. 3

[34] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022. 1, 3, 6, 7, 8

[35] M Zaigham Zaheer, Arif Mahmood, M Haris Khan, Mattia Segu, Fisher Yu, and Seung-Ik Lee. Generative cooperative learning for unsupervised video anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14744–14754, 2022. 3

[36] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 3, 6

[37] Jiangning Zhang, Xuhai Chen, Yabiao Wang, Chengjie Wang, Yong Liu, Xiangtai Li, Ming-Hsuan Yang, and Dacheng Tao. Exploring plain vit reconstruction for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2312.07495*, 2023. 3

[38] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3914–3923, 2023. 1, 6, 7, 8

[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. *arXiv preprint arXiv:2310.07716*, 2023. 3

[40] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pages 392–408. Springer, 2022. 1, 2, 3, 4, 6, 7