PAD: A Dataset and Benchmark for Pose-agnostic Anomaly Detection

Qiang Zhou[1*] **Weize Li**[1*†] **Lihan Jiang**[2 †] **Guoliang Wang**[1†]
Guyue Zhou[1] **Shanghang Zhang**[3] **Hao Zhao**[1]
[1]AIR, Tsinghua University [2]Wuhan University [3]National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
{bamboosdu, liweize0224}@gmail.com shanghang@pku.edu.cn
zhaohao@air.tsinghua.edu.cn

Abstract

Object anomaly detection is an important problem in the field of machine vision and has seen remarkable progress recently. However, two significant challenges hinder its research and application. First, existing datasets lack comprehensive visual information from various pose angles. They usually have an unrealistic assumption that the anomaly-free training dataset is pose-aligned, and the testing samples have the same pose as the training data. However, in practice, anomaly may exist in any regions on a object, the training and query samples may have different poses, calling for the study on pose-agnostic anomaly detection. Second, the absence of a consensus on experimental protocols for pose-agnostic anomaly detection leads to unfair comparisons of different methods, hindering the research on pose-agnostic anomaly detection. To address these issues, we develop Multi-pose Anomaly Detection (MAD) dataset and Pose-agnostic Anomaly Detection (PAD) benchmark, which takes the first step to address the pose-agnostic anomaly detection problem. Specifically, we build MAD using 20 complex-shaped LEGO toys including 4K views with various poses, and high-quality and diverse 3D anomalies in both simulated and real environments. Additionally, we propose a novel method OmniposeAD, trained using MAD, specifically designed for pose-agnostic anomaly detection. Through comprehensive evaluations, we demonstrate the relevance of our dataset and method. Furthermore, we provide an open-source benchmark library, including dataset and baseline methods that cover 8 anomaly detection paradigms, to facilitate future research and application in this domain. Code, data, and models are publicly available at https://github.com/EricLee0224/PAD.

1 Introduction

Unsupervised visual anomaly detection attempts to detect shape or texture anomalies using normal samples and representations from pre-trained model. Due to the increasing demand for quality control of manufactured products and the complicated shape of object and its texture information, object-centric anomaly detection has attracted growing interest. However, existing object anomaly detection datasets and methods are designed under the pose-aligned assumption, which **limits the ability to detect potentially occluded anomalies from arbitrary pose views.** By overcoming the inherent constraints of pose-aligned anomaly detection approaches, the pose-agnostic anomaly detection setting offers enhanced flexibility and applicability in real-world scenarios where a object may exhibit anomalies from diverse perspectives.

[*]Equal contribution.
[†]Work done during internship at AIR.

37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks.

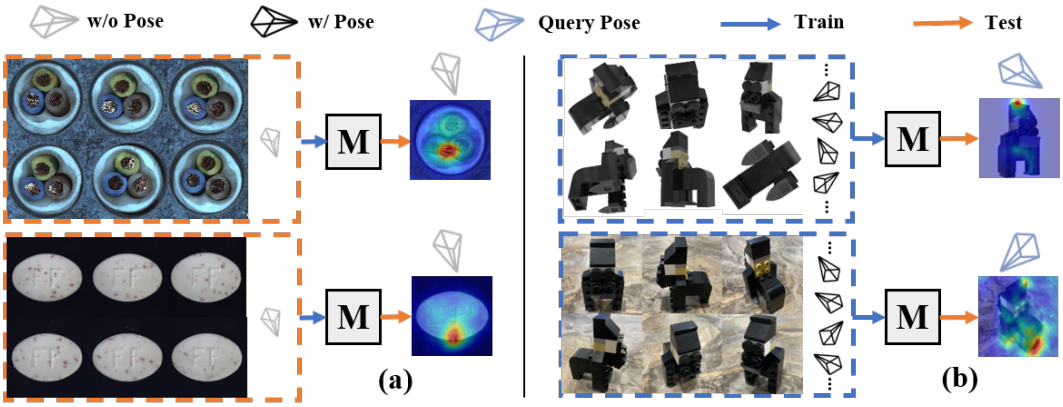

Figure 1: (a) Existing assumption for Pose-aligned Anomaly Detection; (b) Our introduced Pose-agnostic Anomaly Detection setting. **M** denotes pre-trained AD model.

To address the pose-agnostic anomaly detection problem, two key challenges need to be overcome. First, we currently lack datasets that simultaneously provide RGB images of object's multiple viewpoints and its pose labels. This limitation hinders the model's ability to learn the decision boundaries in high-level feature space of anomaly-free object in terms of both shape and texture, which is critical for performing accurate pose-agnostic anomaly detection. Second, in addition to above mentioned dataset being needed, how to measure and compare the performance of pose-agnostic anomaly detection models remains an issue. Existing benchmarks focus on pose-aligned anomaly detection, making it difficult to establish fair criterion for evaluating the effectiveness of pose-agnostic anomaly detection methods.

To this end, we developed the multi-pose anomaly detection (MAD) dataset that contributes to a comprehensive evaluation of pose-agnostic anomaly detection. Our dataset consists of 20 diverse LEGO toys, capturing more than 11,000 images from different viewpoints covering a wide range of poses. Furthermore, the dataset includes three types of anomalies that was carefully selected to reflect both the simulated and real-world environments. To explore our MAD dataset, we make in-depth analyses from two perspectives (shape and texture) with interesting observations. We conducted the analysis of the shape and texture complexity for different categories of LEGO toys in the dataset to gain insight into the performance differences of the baseline methods. By investigating the correlation between model performance and attribute complexity, we aim to assess whether the baseline methods are effective in capturing the full range of visual information from normal objects. Alongside the dataset, we propose OmniposeAD, which introduces neural radiance field into the anomaly detection paradigm for the first time. It overcomes the difficulty of learning the decision boundaries of a normal object from both shape and texture, has the ability to detect anomalies in multiple views of an object, and is even able to detect anomalies in poses that have never been seen in the training set. The main contributions are summarized as follows:

- We introduced **P**ose-agnostic **A**nomaly **D**etection (PAD), a challenging setting for object anomaly detection from diverse viewpoints/poses, breaking free from the constraints of stereotypical pose alignment, and taking a step forward to pratical anomaly detection and localization tasks.
- We developed the **M**ulti-pose **A**nomaly **D**etection (MAD) dataset, the first attempt at evaluating the pose-agnostic anomaly detection. Our dataset comprises MAD-Sim and MAD-Real, containing 11,000+ high-resolution RGB images of multi-pose views from 20 diverse shapes and colors of LEGO toys, offering pixel-precise ground truth for 3 types of anomalies.
- We conducted a comprehensive benchmark of the PAD setting, utilizing the MAD dataset. Across 8 distinct paradigms, we meticulously selected 10 state-of-the-art methods for evaluation. In a pioneering effort, we delving into the previously unexplored realm of correlation between object attributes and anomaly detection performance.
- We proposed ***OmniposeAD***, utilizing NeRF to encode anomaly-free objects from diverse viewpoints/poses and comparing reconstructed normal reference with query image for pose-agnostic anomaly detection and localization. It outperforms previous methods quantitatively and qualitatively on the MAD benchmark, which indicates the promises of PAD setting.

Table 1: Comparison of MAD and existing object anomaly detection datasets. Abbreviated letters refer to the details[1].

Datasets	Year.	Type.	Represent.	Sample statistics			Object Attributes		
				#Cls.	#Normal	#Abnormal	Color	Structure	Pose
GDXray[25]	2015	Real	RGB-D	1	0	19,407	Grayscale	C	✗
MVTec:									
AD[5]	2019	Real	RGB	15	4,096	1,258	Diverse	S	✗
3D AD[7]	2021	Real	RGB/PC	15	2,904	948	Diverse	S	✗
LOCO-AD[4]	2022	Real	RGB	5	2,347	993	Diverse	C	✗
MPDD[20]	2021	Real	RGB	6	1,064	282	Diverse	C	✗
Eyecandies[8]	2022	Syn.	RGB/D/N	10	13,250	2,250	Diverse	S	✗
VisA[65]	2022	Real	RGB	12	9,621	1,200	Diverse	S&C	✗
MAD (Ours)	2023	Sim+Real	RGB	20	5,231	4,902	Diverse	S&C	✓

2 Related Work

Object Anomaly Detection Datasets. The progress of object anomaly detection in industrial vision is significantly impeded by the scarcity of datasets containing high-quality annotated anomaly samples and comprehensive view information about normal objects. MVTec has developed a series of widely-used photo-realistic industrial anomaly detection datasets: The objects provided by the AD[5] dataset are simple, as discerning anomalies can be achieved solely from a single view. Although the 3D-AD[7] dataset offers more complex objects, it lacks RGB information from a full range of views, requiring the supplementation of hard-to-capture point cloud data to detect subtle structural anomalies. The LOCO AD[4] dataset provides rich global structural and logical information but is not suitable for fine-grained anomaly detection on individual objects. GDXray[25] provides grayscale maps obtained through X-ray scans for visual discrimination of structural defects but lacks normal samples and color/texture information. The MPDD[20] dataset offers multi-angle information about the objects but is limited in size and lacks standardized backgrounds in the photos. Recently, Eyecandies[8] has introduced a substantial collection of synthetic candy views captured under various lighting conditions and provides multi sensory inputs. However, there remains a significant gap between synthesized data and the real or simulated data domain. VisA[65] contains objects with complex structures, multiple instances at different locations in single view, different objects across 3 domains, and multiple anomaly classes, thus is closer to the real world than MvTec-AD[5]. To address these issues and enable exploration of the pose-agnostic AD problem, we propose Multi-pose Anomaly Detection (MAD) dataset. We present comprehensive comparison between MAD and other representative object anomaly detection datasets in Table.1.

Unsupervised Anomaly Detection for Visual Inspection. Existing methods can be categorized into feature embedding-based and reconstruction-based strategies, according to a recent survey[24]. Feature embedding-based methods such as Teacher-Student Architecture[6, 51, 15, 10, 35], One-Class Classification[17, 56, 62, 31, 22, 54], Distribution-map[36, 42, 33, 34, 18, 59, 44], and Memory Bank[32, 13, 21, 2, 50] aim to learn low-dimensional representations of input data to capture underlying patterns and anomalies. Reconstruction-based methods exploiting Autoencoder (AE)[3, 53, 61, 60, 14, 46], Generative Adversarial Networks (GANs)[38, 37, 52, 39, 23], and Transformer[27, 58, 29, 57] model aim to learn a representation that can effectively reconstruct normal samples. Deviations in reconstruction quality indicate the presence of anomalies. BTF[19] reveals that using only angularly limited RGB images to determine whether an object is defective or not can easily overlook some obscured structural anomalies. Therefore it is necessary to develop methods for 3D anomaly detection using 3D information, such as point clouds, depth, etc. AST[35] employs RGB image with depth information to enhance anomaly detection performance. M3DM[47] and CPMF[11] encourage the fusion of features from different modalities in RGB and point cloud information, and combine the local geometric information of the 3D modal with the global semantic information of the pseudo 2D modal.

Neural Radiance Field. Neural Radiation Field (NeRF)[26] implements an implicit representation of the scene, receiving hundreds of sampling points along each camera ray and outputting the predicted color and density. Because NeRF can handle complex illumination and occlusion situations, we can use it to capture the color and shape properties of objects in real scenes, which means that NeRF has

[1]Syn., PC, D, N, C and S denotes Synthesis, Point Cloud, Depth, Normal Map, Complex, and Simple, respectively.

the potential to be applied to anomaly detection tasks. It has also been used for numerous applications, such as large-scale urban reconstruction[48, 49, 40], human face or body reconstruction[55, 12], and robotic applications such as pose estimation and SLAM[41, 64, 63]. Since NeRF allows for the continuous modeling of 3D scenes and the re-rendering of high-quality photos, it can create reference images from any angle for anomaly detection in this work.

3 MAD: Multi-pose Anomaly Detection Dataset

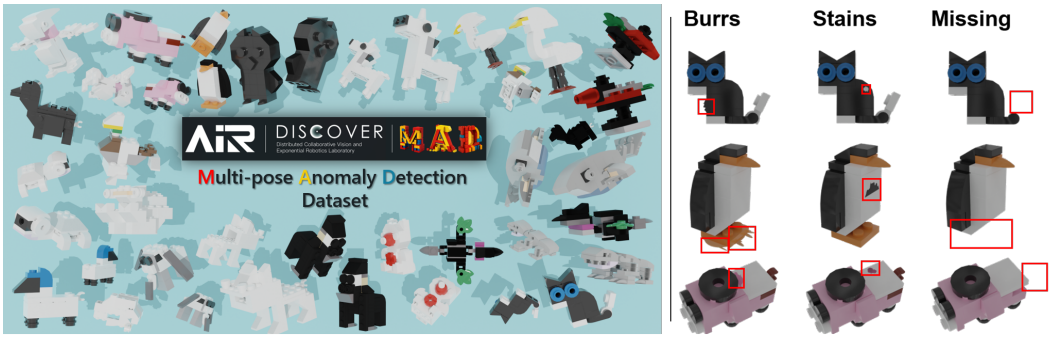

Figure 2: (a) Our LEGO family in MAD dataset; (b) Examples of three defect types.

3.1 Data Selection

Aiming to support pose-agnostic anomaly detection, our dataset selects 20 LEGO animals toys with diverse shape complexity and color contrasts as target objects shown in Figure.2(a). These selected toys encompass a wide range of shapes and sizes, ranging from *cat* and *pig* to larger animals like *unicorn* and *elephant*, along with other interesting animal forms. Each animal toy is meticulously designed and constructed to ensure a variety of shapes and complexities. In selecting these objects, we also paid attention to color contrasts. Bright and vibrant colors are used in the construction of each animal toy, making them more visually striking and easily recognizable in the images. The choice of color contrasts aims to assist algorithms in accurately detecting and localizing anomalies in different environments and backgrounds.

3.2 Data Processing and Annotation

We used Blender in combination with Ldrew (LEGO parts library) to build 20 different types of 3D LEGO animal models and generate anomalous such as burrs, stains and missing parts shown as Figure.2(b). To achieve a full pose angle rendering of the normal samples, we established the center of the animal as the origin and positioned the rendering camera around the sphere. By maintaining a fixed radius, we generated multiple cameras simultaneously, each positioned at a different theta/phi coordinate in the spherical coordinate system. In addition, we created ten cameras equally spaced along the Z-axis, with the camera orientation set to face the animal. In order to edit the real anomalies on the simulation model, we hired professional LEGO players to manually modify the 20 types of LEGO animal models according to the types of anomalies encountered using PhotoShop. For the hand-edited anomalies, synchronized ground truth labels were produced by a computer vision researcher. Notably, only a subset of images showed the anomaly portion, which required a manual selection process to determine the appropriate data for anomaly detection. Ultimately, approximately150-300 anomaly images were generated for each animal. To generate training data with more normal instances, we used the same approach. For the real dataset, we collect parts from the actual production line, including abnormal parts characterized by real burrs and stains. In addition, we manually generated the missing-piece anomaly by removing LEGO parts. Given the rarity of anomalies in the manufacturing process of LEGO toys and the difficulty in obtaining these parts, we deliberately reduced the sample size of the real-world version.

3.3 Data Statistics and Split

We present the statistical information of the dataset in Table.1. During the data spliting process, we followed design strategy of existing anomaly detection datasets. MAD-Sim is a dataset created programmatically using Blender, allowing for controlled sample quantities. We divided it into training and test sets with a sample ratio of approximately 2:1 to explore anomaly detection algorithm performance on multi-view objects. For MAD-Real, the data was captured in real-world scenes using a camera. Anomalies are infrequent in actual production, and we aimed to reflect this reality. MAD-Real has approximately one-fifth the total sample count of MAD-Sim and follows a similar division into training and test sets, maintaining a sample ratio of about 4:1.

3.4 Object Attributes Quantification

Anomaly detection algorithms are evolving rapidly, but its often difficult to intuitively determine which one is the best from performance metrics when trying to deploy an algorithm to a real scenario. To reduce testing costs, we explore a novel aspect in anomaly detection tasks, which is the relationship between object attributes and anomaly detection performance. When we are given an object to be tested, we expect to determine which algorithm is likely to be the best, simply by referring to the correlation between quantitative metrics of the object's attributes, i.e., object shape complexity and color contrast, and the performance of different algorithms. The quantitative details are shown in Fig.3.

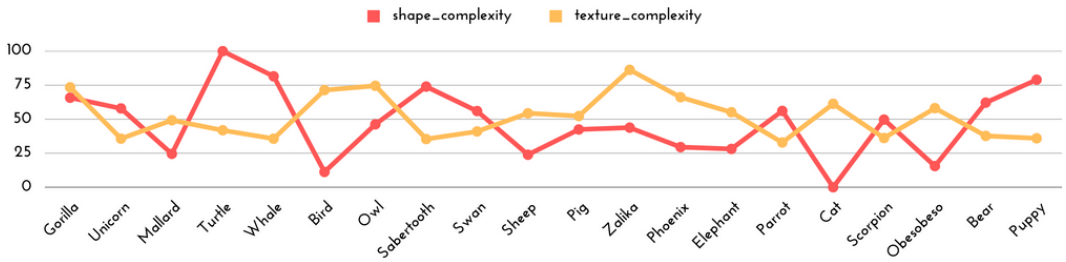

Figure 3: Object attributes quantification of 20 LEGO toys.

4 Pose-agnostic Anomaly Detection Framework

4.1 Problem Definition

Our Pose-agnostic Anomaly Detection (PAD) setting introduced for object anomaly detection and localization tasks is shown as Figure.1 and it can be formally 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 from object's multi pose view and each sample t consists of a RGB image I_{rgb} w/ pose information θ_{pose}. In addition, $\mathcal{T}$ belongs to certain object o_j, $o_j \in \mathcal{O}$, where $\mathcal{O}$ denotes the set of all objects categories. During testing, given a query (normal or abnormal) image $\mathcal{Q}$ from object o_j w/o pose information θ_{pose}, the pre-trained AD model M should discriminate whether or not the query image $\mathcal{Q}$ is anomalous and localize the pixel-wise anomaly region if the anomaly is detected.

4.2 OmniposeAD

The OmniposeAD illustrated in Figure.4, consists of anomaly-free neural radiance field, coarse-to-fine pose estimation module, and anomaly detection and localization module. The input of OmniposeAD is query image $\mathcal{Q}$ w/o pose. Initially, $\mathcal{Q}$ undergoes the coarse-to-fine pose estimation module to obtain the accurate camera view pose $\hat{\theta}_{\text{pose}}$. Subsequently, $\hat{\theta}_{\text{pose}}$ is utilized in the pre-trained neural radiance field for rendering the normal reference. Finally, the reference is compared to the $\mathcal{Q}$ to extract the anomaly information.

4.2.1 Anomaly-free reference view synthesis

To reconstruct anomaly-free references from various viewpoints, we employ unsupervised algorithm capable of synthesizing novel views. With an anomaly-free training set $\mathcal{T}$, we represent anomaly-free

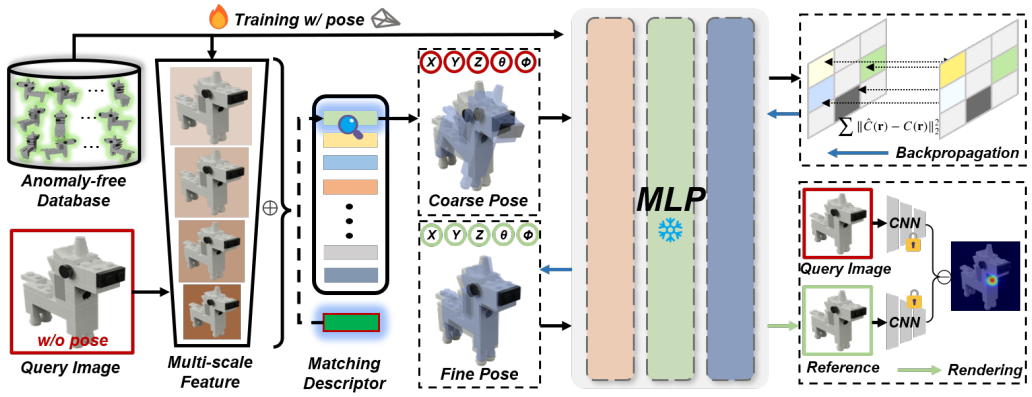

Figure 4: OmniposeAD consists of three componets. 1) The Neural Radiance Field is trained using multi-pose views of anomaly-free objects to capture normal information. 2) The query image proceeds through two-stage pose estimation process, from coarse to fine: a matching descriptor is generated to match with the anomaly-free database samples to obtain coarse pose, then the pose refined using the iNeRF[55] to obtain the reference pose. 3) The pre-trained Neural Radiance Field renders the normal reference, which is compared with the query image for anomaly detection and localization.

objects implicitly using vanilla neural radiance field, following [26] which employs a multi layer perceptron (MLP) network and minimizes the photometric loss.

4.2.2 Coarse-to-fine pose estimation

To generate normal reference using the pre-trained neural radiance field, the fine camera pose $\hat{\theta}_{\text{pose}}$ of the query image needs to be determined. For this purpose, we employ the coarse-to-fine pose estimation module that combines image retrieval techniques and iNeRF[55].

On a coarse scale Our objective is to roughly align the query image $\mathcal{Q}$ with an anomaly-free image with camera pose in the training database i.e., find a function ϕ that satisfies the following criteria:

$$\phi : \mathbb{R}^{H\times W\times 3} \mapsto \mathbb{R}^{64} \quad \phi(\mathcal{Q}) = \mathbf{q} \in \mathbb{R}^{64} \tag{1}$$

First, we utilize the pre-trained EfficientNet-B4 model[43] as the backbone for feature extraction. Following the standard procedure[58], we input the multi-scale feature map into the network to obtain the descriptor $\mathbf{q}$. Given an 3-channel RGB query image I_{rgb} of dimension $H \times W$, our model ϕ extract a compact 64-dimensional ($64\mathcal{D}$) descriptor or embedding $\phi(x)$. Then the image retrieval task[9, 1, 30] considers a database of images I_{rgb} for each sample t_i in the training set $\mathcal{T}$, and given query image $\mathcal{Q}$, we calculate:

$$\underset{\mathcal{T}}{\operatorname{argmin}} \|\phi(\mathcal{Q}) - \phi(I)\|_2^2 \tag{2}$$

Finally, the top-1 most similar image I_{rgb} is retrieved. We utilize the associated pose θ_{pose} of the image I_{rgb} as the initial coarse pose $\hat{\theta}_{\text{pose}}$ for the next stage.

On a fine scale We refine the estimated pose using iNeRF[55]. This method enables us to fully utilize the pre-trained NeRF model and estimate camera poses. iNeRF uses the ability from NeRF to take some estimated camera pose $\theta \in \text{SE(3)}$ in the coordinate frame of the NeRF model and render a corresponding image observation, and then use the same photometric loss function as NeRF but updates the fine pose $\hat{\theta}_{\text{pose}}$ instead of the weights of the MLP network.

4.2.3 Anomaly detection and localization

When the normal reference is rendered using $\hat{\theta}_{\text{pose}}$ and query image $\mathcal{Q}$ are available, we use the frozen pre-trained CNN backbone for feature extraction. The feature from layer1 to layer5 are resized

to the same size as the query image. Then the multi-scale feature and the query image are concatenate together to form a feature map, $q \in R^{C\times H\times W}$ with the corresponding input pixel. The query image corresponds to $f_q \in R^{C\times H\times W}$, while the rendered reference corresponds to $\hat{f} \in R^{C\times H\times W}$. We first define the feature difference map, $d(i,u)$, as shown in Eqn. 3

$$\boldsymbol{d}(i,u) = \boldsymbol{f_q}(i,u) - \hat{\boldsymbol{f}}(i,u), \tag{3}$$

where i denotes the channel index and, u is the spatial position index (height together with width for simplicity).

$$\boldsymbol{s}(u) = ||\boldsymbol{d}(:,u)||_2. \tag{4}$$

Anomaly detection aims to determine if an image contains anomalous regions. We intuitively use the maximum value of the averagely pooled $\boldsymbol{s}(u)$ as the anomaly score of the whole image.

Anomaly localization aims to localize anomalous regions, producing an anomaly score map, $\boldsymbol{s}(u)$ in Eqn. 4 , which assigns an anomaly score for each pixel, u. $\boldsymbol{s}(u)$ is calculated as the $L2$ norm of the feature difference vector, $\boldsymbol{d}(:,u)$ in Eqn. 3.

Table 2: Quantitative results of pixel/image-level anomaly detection performance on MAD.

Category	Feature Embedding-based						Reconstruction-based				Ours
	Patchcore[32]	STFPM[45]	Fastflow[59]	CFlow[18]	CFA[21]	Cutpaste[22]	DRAEM[16]	FAVAE[51]	OCRGAN[28]	UniAD[57]	
Gorilla	88.4/66.8	93.8/65.3	91.4/51.1	94.7/69.2	91.4/41.8	36.1/-	77.7/58.9	92.1/46.8	94.2/-	93.4/56.6	**99.5/93.6**
Unicorn	58.9/92.4	89.3/79.6	77.9/45.0	89.9/82.3	85.2/85.6	69.6/-	26.0/70.4	88.0/68.3	86.7/-	86.8/73.0	**98.2/94.0**
Mallard	66.1/59.3	86.0/42.2	85.0/72.1	87.3/74.9	83.7/36.6	40.9/-	47.8/34.5	85.3/33.6	88.9/-	85.4/70.0	**97.4/84.7**
Turtle	77.5/87.0	91.0/64.4	83.9/67.7	90.2/51.0	88.7/58.3	77.2/-	45.3/18.4	89.9/82.8	76.7/-	88.9/50.2	**99.1/95.6**
Whale	60.9/**86.0**	88.6/64.1	86.5/53.2	89.2/57.0	87.9/77.7	66.8/-	55.9/65.8	90.1/62.5	89.4/-	90.7/75.5	**98.3**/82.5
Bird	88.6/82.9	90.6/52.4	90.4/76.5	91.8/75.6	92.2/78.4	71.7/-	60.3/69.1	91.6/73.3	**99.1**/-	91.1/74.7	95.7/**92.4**
Owl	86.3/72.9	91.8/72.7	90.7/58.2	94.6/76.5	93.9/74.0	51.9/-	78.9/67.2	96.7/62.5	90.1/-	92.8/65.3	**99.4/88.2**
Sabertooth	69.4/76.6	89.3/56.0	88.7/70.5	93.3/71.3	88.0/64.2	71.2/-	26.2/68.6	94.5/82.4	91.7/-	90.3/61.2	**98.5/95.7**
Swan	73.5/75.2	90.8/53.6	89.5/63.9	93.1/67.4	95.0/66.7	57.2/-	75.9/59.7	87.4/50.6	72.2/-	90.6/57.5	**98.8/86.5**
Sheep	79.9/89.4	93.2/56.5	91.0/71.4	94.3/80.9	94.1/86.5	67.2/-	70.5/59.5	94.3/74.9	**98.9**/-	92.9/70.4	97.7/**90.1**
Pig	83.5/85.7	94.2/50.6	93.6/59.6	97.1/72.1	95.6/66.7	52.3/-	65.6/64.4	92.2/52.5	93.6/-	94.8/54.6	**97.7/88.3**
Zalika	64.9/68.2	86.2/53.7	84.6/54.9	89.4/66.9	87.7/52.1	43.5/-	66.6/51.7	86.4/34.6	94.4/-	86.7/50.5	**99.1/88.2**
Phoenix	62.4/71.4	86.1/56.7	85.7/53.4	87.3/64.4	87.0/65.9	53.1/-	38.7/53.1	92.4/65.2	86.8/-	84.7/55.4	**99.4/82.3**
Elephant	56.2/78.6	76.8/61.7	76.8/61.6	72.4/70.1	77.8/71.7	56.9/-	55.9/62.5	72.0/49.1	91.7/-	70.7/59.3	**99.0/92.5**
Parrot	70.7/78.0	84.0/61.1	84.0/53.4	86.8/67.9	83.7/69.8	55.4/-	34.4/62.3	87.7/46.1	66.5/-	85.6/53.4	**99.5/97.0**
Cat	85.6/78.7	93.7/52.2	93.7/51.3	94.7/65.8	95.0/68.2	58.3/-	79.4/61.3	94.0/53.2	91.3/-	93.8/53.1	**97.7/84.9**
Scorpion	79.9/82.1	90.7/68.9	74.3/51.9	91.9/79.5	92.2/91.4	71.2/-	79.7/83.7	88.4/66.9	**97.6**/-	92.2/69.5	95.9/**91.5**
Obesobeso	91.9/89.5	94.2/60.8	92.9/67.6	95.8/80.0	96.2/80.6	73.3/-	89.2/73.9	92.7/58.2	**98.5**/-	93.6/67.7	98.0/**97.1**
Bear	79.5/84.2	90.6/60.7	85.0/72.9	92.2/81.4	90.7/78.7	68.8/-	39.2/76.1	90.1/52.8	83.1/-	90.9/65.1	**99.3/98.8**
Puppy	73.3/65.6	84.9/56.7	80.3/59.5	89.6/71.4	82.3/53.7	43.2/-	45.8/57.4	85.6/43.5	78.9/-	87.1/55.6	**98.8/93.5**
Mean	74.7/78.5	89.3/59.5	86.1/60.8	90.8/71.3	89.8/68.2	59.3/-	58.0/60.9	89.4/58.0	88.5/-	89.1/62.2	**97.8/90.9**

5 Experiments and Analysis

Our experiments aim to broadly evaluate the performance of anomaly detection algorithms in a pose-agnostic setting, where different types of anomalies may appear in different object poses. We selected representative state-of-the-art methods from different paradigms for fair preliminary benchmarking using the MAD dataset we developed with the aim of exploring the applicability of existing methods in the PAD setting. In addition, we evaluate the performance of our proposed OmniposeAD, normal reference reconstruction results, and the effect of training poses/viewpoints richness on performance.

Benchmark Methods Selection. The selection criteria for benchmark methods include representativeness, superior performance, and availability of source code. To comprehensively investigate the performance of anomaly detection algorithms in the pose-agnostic anomaly detection setting, we selected representative methods from 8 paradigms, the taxonomy can be found in the supplementary. For the Feature Embedding-based strategy, we selected Patchcore[32], STFPM[45], Fastflow[59], CFlow[18], CFA[21], and Cutpaste[22]. For the Reconstruction-based strategy, we selected DRAEM[16], FAVAE[51], OCR-GAN[28], and UniAD[57]. It is important to note that 3D anomaly detection methods utilizing multimodal information were not considered in order to ensure a fair comparison, as the MAD dataset only provides RGB information.

Implementation Details. For the implementation details, we trained the OmniposeAD using the MAD dataset, employing the same hyperparameters for both synthetic and real scenes. During the anomaly-free novel view synthesis phase, the process took approximately 5 hours with a batch size of 4096. The image resolution was set to 400x400 for the simulated dataset and 504x378 for the

real dataset to ensure efficient processing. Additionally, the coarse-to-fine pose estimation process involved a total of 300 optimization steps. We initialized the learning rate to 0.01 and decayed it to 5.8e-05 at epoch 50 to facilitate a smooth optimization process, utilizing a batch size of 3072. The entire framework required approximately 10-15 hours to run on a single NVIDIA Tesla A100.

Evaluation Metric. Following previous work, we specifically choose the Area Under the Receiver Operating Caracteristic Curve (AUROC) as the primary metric for evaluating the performance of anomaly segmentation at the pixel-level and anomaly classification at the image-level. While there exist various evaluation metric for these tasks, AUROC stands out as the most widely used and suitable metric for conducting comprehensive benchmarking. The AUROC score can be calculated as follows:

$$\text{AUROC} = \int (\text{R}_{\text{TP}}) \text{dR}_{\text{FP}} \tag{5}$$

Here, R_{TP} and R_{FP} represent the pixel/image-level true positive rate and false positive rate, respectively.

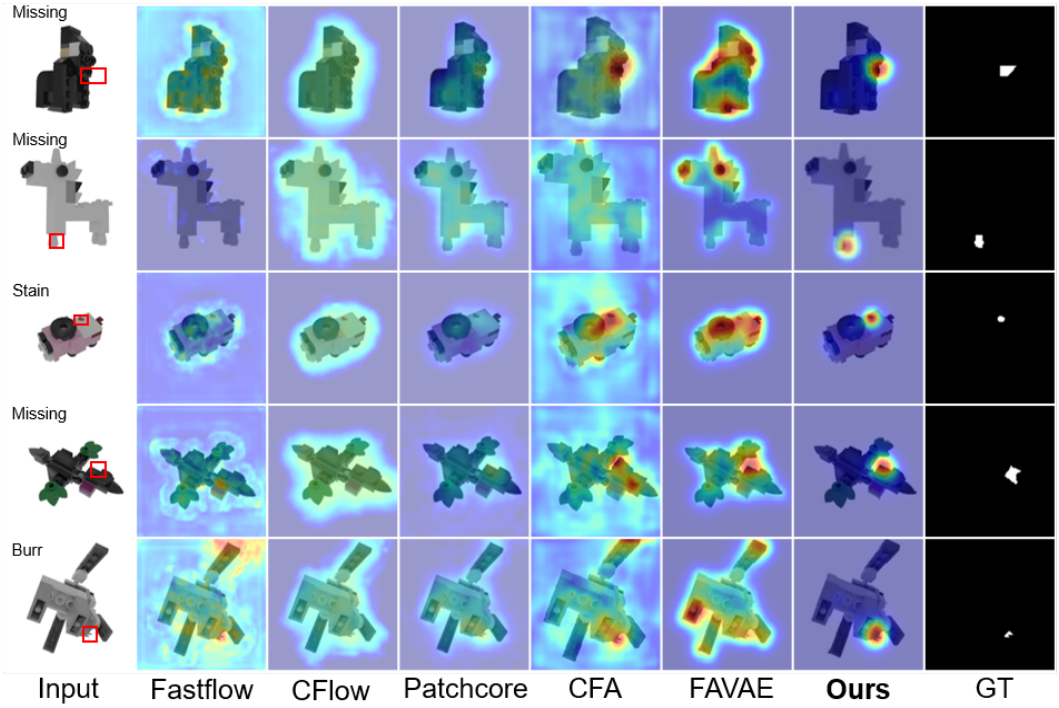

Figure 5: Qualitative results visualization of anomaly localization performance on MAD-Sim.

5.1 Pose-agnostic Anomaly Detection Benchmark on MAD Dataset

Our experiments provide a fair comparison between the state-of-the-art methods and OmniposeAD in the PAD setting, including pixel and image-level AUROC performance. Table 2 and Figure 5 clearly demonstrate that OmniposeAD achieves the state-of-the-art performance on MAD dataset for pose-agnostic anomaly detection evaluation. Specifically, OmniposeAD outperforms Feature Embedding-based CFlow[18] and Reconstruction-based UniAD[57] by a large margin, +7.0/+19.6 and +8.7/+28.7, respectively. Its noteworthy that additional data (i.e., using a pre-trained model on ImageNet during the training phase) always brings advantage. This consistent superiority can be attributed to ImageNet's extensive repository of object representations spanning diverse poses. In a direct comparison with methods not utilizing additional training data, our approach showcases a remarkable enhancement, as evidenced by a substantial increment of +29.2/+8.3. Both the Feature Embedding and Reconstruction-based methods, namely DRAEM[60] and Cutpaste[22], employ the generation of pseudo-anomalies as a data augmentation technique. However, these approaches

exhibit a notable decline in performance within the PAD setting. Objects may present different representations in different poses, leading to complex patterns of visual anomalies. It is difficult to effectively characterize the anomalies of objects under different viewpoints through simple data enhancement methods.

Our benchmarks demonstrate the considerable potential of both feature embedding-based and reconstruction-based approaches in the PAD setting. In particular, for the pixel-level anomaly localization task, the performance of STFPM[45](89.3), CFlow[18](90.8), CFA[21](89.8), FAVAE[51](89.4), and UniAD[57](89.1) shows a slightly fluctuating but largely respectable performance. However, with the exception of our proposed OmniposeAD, the performance of benchmark methods in image-level anomaly detection tasks is not satisfactory. It is worth noting that UniAD[57], although capable of learning multi-category anomaly boundaries using a single model, still struggles to achieve satisfactory image-level anomaly detection performance for multi-pose anomalies in the PAD setting. To address this gap, algorithms that are adept at capturing the normal attributes of multi-pose objects need to be explored to achieve reasonable image-level anomaly detection performance within the PAD setting.

5.2 Correlation of performance with object attributes

The results are shown in Figure.6, where the performance of most methods is positively correlated with color contrast and negatively correlated with structure complexity, which is consistent with our intuition. Notably, Cutpaste[22], a representative approach that generates anomalies and reconstructs them through a self-supervised task, stands out as being sensitive to color contrast but surprisingly tolerant towards shape complexity. Furthermore, we demonstrate the robustness of our proposed OmniposeAD to changes in object attributes.

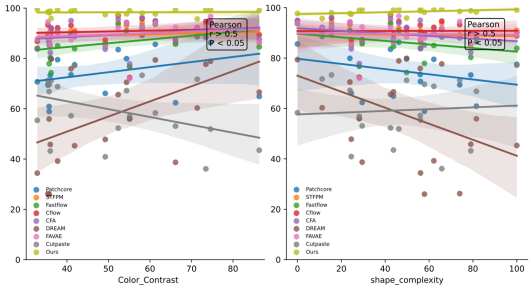

Figure 6: Correlation between object attributes and anomaly detection performance.

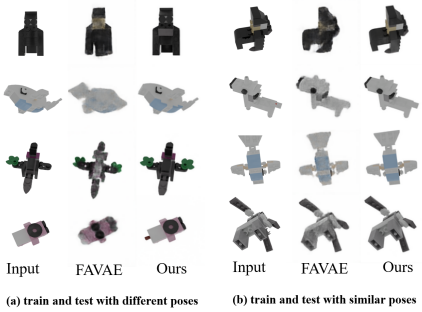

Figure 7: Visualization of reference reconstruction between OmniposeAD and FAVAE.

5.3 Validation of reference reconstruction on *OmniposeAD*

Both the reconstruction-based method and our proposed OmniposeAD essentially reconstruct the normal reference. In the quantitative results of the anomaly localization task, the representative reconstruction-based method FAVAE achieved 89.4, and we further visualized the reconstructed references to explore the gap. Figure.7 shows that FAVAE does not reconstruct references for unseen poses well (the quantitative results do not align with visual facts), hinting that the existing evaluation metrics are potentially unsatisfactory in the PAD setting. Our further discussion is provided in Sec.4 of the Supplementary.

5.4 Effect of Dense-to-Sparse training data on *OmniposeAD* performance

In this section, we utilize the MAD-Real dataset to evaluate the impact of the transition from dense to sparse viewpoint training data on the performance of the OmniposeAD method, reflecting the lack of available viewpoints for real-world applications. We control the sample size of the training set for the OmniposeAD model to be 100, 70, and 50, respectively. As shown in Table 3: Despite the sparse training data leads to a slight decrease in performance, the less pronounced drop in pixel-

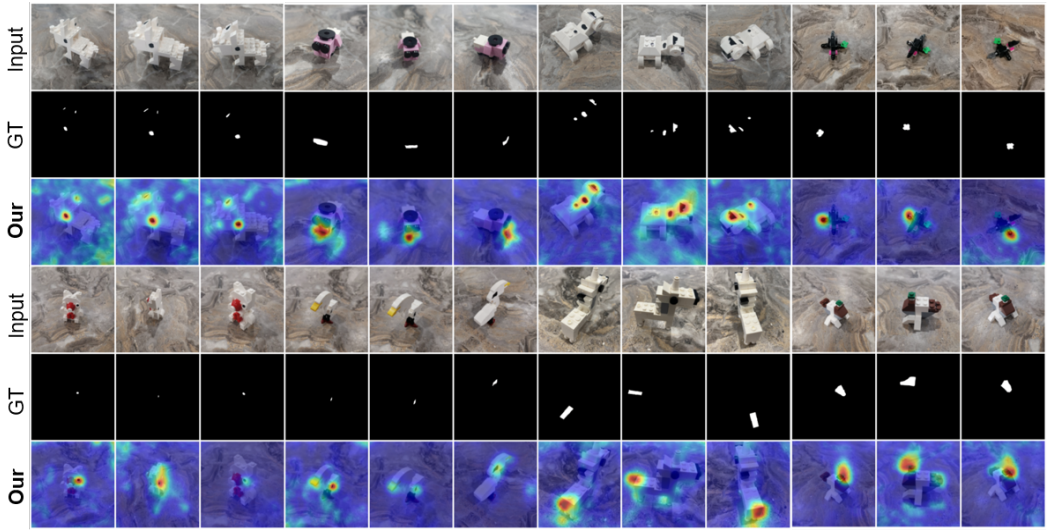

Figure 8: Qualitative results visualization of OmniposeAD anomaly localization performance from multi-view on MAD-Real.

level performance suggests that OmniposeAD excels in capturing local consistency within objects. The more significant decrease at the image-level indicates that sparse data impacts OmniposeAD's ability to learn the holistic normal features of object from global perspective. We further showed the multi-view anomaly localization performance of OmniposeAD on the MAD-Real dataset in Figure.8.

Table 3: Dense-view to sparse-view object anomaly detection results on MAD-Real.

Sample Size	100		70		50		Sample Size	100		70		50	
Object Class	Image	Pixel	Image	Pixel	Image	Pixel	Object Class	Image	Pixel	Image	Pixel	Image	Pixel
Gorilla	**93.6**	**99.5**	85.4	97.5	91.0	97.4	Pig	**75.7**	**96.4**	73.9	95.7	72.8	95.4
Unicorn	**86.2**	**96.5**	84.4	96.1	80.2	94.9	Zalika	**88.2**	**99.1**	78.8	96.8	72.8	96.0
Mallard	**87.3**	**96.3**	80.3	94.4	74.6	95.1	Phoenix	**86.0**	**99.4**	84.6	99.2	82.6	98.8
Turtle	**99.1**	**95.4**	95.1	92.3	83.4	90.4	Elephant	**91.0**	**98.1**	82.5	95.9	85.1	96.5
Whale	**82.1**	**98.5**	82.5	97.0	76.0	93.1	Parrot	**91.5**	**99.1**	85.5	98.2	87.9	96.9
Bird	**92.0**	**95.1**	91.6	91.7	89.1	93.4	Cat	**78.9**	**97.6**	73.5	97.6	73.7	97.1
Owl	**89.1**	**99.3**	78.8	98.1	89.1	98.8	Scorpion	**87.4**	**94.2**	77.0	91.8	74.3	91.5
Sabertooth	**93.8**	**97.8**	84.9	95.1	81.5	94.4	Obesobeso	**88.5**	**96.1**	86.4	97.9	81.8	97.8
Swan	**80.1**	**98.0**	77.0	97.3	71.3	96.7	Bear	**97.8**	**99.3**	92.0	98.0	87.8	97.9
Sheep	**84.3**	**97.6**	83.2	97.6	82.3	92.8	Puppy	**93.1**	**98.5**	86.5	95.8	80.3	95.3

6 Discussion

Conclusion. In this work, we introduce pose-agnostic anomaly detection setting, and develop the MAD dataset, which is the first exploration to evaluate pose-agnostic anomaly detection algorithms. In addition, we propose OmniposeAD to solve the problem of unsupervised anomaly detection in different poses of an object, which avoids missing potential occluded anomaly regions. **Limitation.** For the MAD dataset, which includes only standard rigid objects and no naturally varying objects, the complexity and diversity of industrial products made it difficult to collect and quantify a complete sample of diverse features. For OmniposeAD, it is difficult to generalize to novel classes to obtain appreciable performance; for all modules, we chose the vanilla method in pursuit of robustness, which has the potential for faster training times, less training data, and more realistic high-frequency details.

7 Acknowledgment

We would like to express our sincere appreciation to Shenzhen Qianzhi Technology Co.,Ltd for their support in providing Lego toys manufacturing and defect samples.

References

[1] André Araujo. Deep image features for instance-level recognition and matching. In *Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents*, pages 1–1, 2020.

[2] Jaehyeok Bae, Jae-Han Lee, and Seyun Kim. Image anomaly detection and localization with position and neighborhood information. *arXiv preprint arXiv:2211.12634*, 2022.

[3] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pages 161–169. Springer, 2019.

[4] 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.

[5] 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.

[6] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020.

[7] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021.

[8] 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 16th Asian Conference on Computer Vision (ACCV2022*, 2022. ACCV.

[9] Bingyi Cao, Andre Araujo, and Jack Sim. Unifying deep local and global features for image search. In *European Conference on Computer Vision*, pages 726–743. Springer, 2020.

[10] Yunkang Cao, Qian Wan, Weiming Shen, and Liang Gao. Informative knowledge distillation for image anomaly segmentation. *Knowledge-Based Systems*, 248:108846, 2022.

[11] Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *arXiv preprint arXiv:2303.13194*, 2023.

[12] Xiaoxue Chen, Junchen Liu, Hao Zhao, Guyue Zhou, and Ya-Qin Zhang. Nerrf: 3d reconstruction and view synthesis for transparent and specular objects with neural refractive-reflective fields. *arXiv preprint arXiv:2309.13039*, 2023.

[13] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020.

[14] David Dehaene and Pierre Eline. Anomaly localization by modeling perceptual features. *arXiv preprint arXiv:2008.05369*, 2020.

[15] 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.

[16] Okwudili M Ezeme, Qusay H Mahmoud, and Akramul Azim. Dream: deep recursive attentive model for anomaly detection in kernel events. *IEEE Access*, 7:18860–18870, 2019.

[17] Max K Ferguson, AK Ronay, Yung-Tsun Tina Lee, and Kincho H Law. Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. *Smart and sustainable manufacturing systems*, 2, 2018.

[18] 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.

[19] Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2967–2976, 2023.

[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.

[21] Sungwook Lee, Seunghyun Lee, and Byung Cheol Song. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *arXiv preprint arXiv:2206.04325*, 2022.

[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.

[23] Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *arXiv preprint arXiv:2203.00259*, 2022.

[24] Jiaqi Liu, Guoyang Xie, Jingbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep industrial image anomaly detection: A survey. *arXiv preprint arXiv:2301.11514*, 2023.

[25] 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*, 34:1–12, 2015.

[26] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[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.

[28] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2898–2906, 2019.

[29] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022.

[30] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1655–1668, 2018.

[31] Tal Reiss, Niv Cohen, Liron Bergman, and Yedid Hoshen. Panda: Adapting pretrained features for anomaly detection and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2806–2814, 2021.

[32] 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.

[33] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but differnet: Semi-supervised defect detection with normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1907–1916, 2021.

[34] 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.

[35] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2592–2602, 2023.

[36] M Sabokrou, M Fayyaz, M Fathy, and R Klette. Fully convolutional neural network for fast anomaly detection in crowded scenes. *arXiv preprint arXiv:1609.00866*, 2016.

[37] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019.

[38] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.

[39] Jouwon Song, Kyeongbo Kong, Ye-In Park, Seong-Gyun Kim, and Suk-Ju Kang. Anoseg: anomaly segmentation network using self-supervised learning. *arXiv preprint arXiv:2110.03396*, 2021.

[40] Hao Su, Charles R Qi, Yangyan Li, and Leonidas J Guibas. Render for cnn: Viewpoint estimation in images using cnns trained with rendered 3d model views. In *Proceedings of the IEEE international conference on computer vision*, pages 2686–2694, 2015.

[41] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6229–6238, 2021.

[42] Matias Tailanian, Pablo Musé, and Álvaro Pardo. A multi-scale a contrario method for unsupervised image anomaly detection. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 179–184. IEEE, 2021.

[43] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[44] Beiwen Tian, Mingdao Liu, Huan-ang Gao, Pengfei Li, Hao Zhao, and Guyue Zhou. Unsupervised road anomaly detection with language anchors. In *2023 IEEE international conference on robotics and automation (ICRA)*, pages 7778–7785. IEEE, 2023.

[45] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for unsupervised anomaly detection. *arXiv preprint arXiv:2103.04257*, 2021.

[46] Lu Wang, Dongkai Zhang, Jiahao Guo, and Yuexing Han. Image anomaly detection using normal data only by latent space resampling. *Applied Sciences*, 10(23):8660, 2020.

[47] 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.

[48] Zirui Wu, Tianyu Liu, Liyi Luo, Zhide Zhong, Jianteng Chen, Hongmin Xiao, Chao Hou, Haozhe Lou, Yuantao Chen, Runyi Yang, et al. Mars: An instance-aware, modular and realistic simulator for autonomous driving. *arXiv preprint arXiv:2307.15058*, 2023.

[49] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. *arXiv preprint arXiv:1711.00199*, 2017.

[50] Guoyang Xie, Jingbao Wang, Jiaqi Liu, Feng Zheng, and Yaochu Jin. Pushing the limits of fewshot anomaly detection in industry vision: Graphcore. *arXiv preprint arXiv:2301.12082*, 2023.

[51] Shinji Yamada and Kazuhiro Hotta. Reconstruction student with attention for student-teacher pyramid matching. *arXiv preprint arXiv:2111.15376*, 2021.

[52] Xudong Yan, Huaidong Zhang, Xuemiao Xu, Xiaowei Hu, and Pheng-Ann Heng. Learning semantic context from normal samples for unsupervised anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3110–3118, 2021.

[53] Yi Yan, Deming Wang, Guangliang Zhou, and Qijun Chen. Unsupervised anomaly segmentation via multilevel image reconstruction and adaptive attention-level transition. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.

[54] 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.

[55] Lin Yen-Chen, Pete Florence, Jonathan T Barron, Alberto Rodriguez, Phillip Isola, and Tsung-Yi Lin. inerf: Inverting neural radiance fields for pose estimation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1323–1330. IEEE, 2021.

[56] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[57] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *arXiv preprint arXiv:2206.03687*, 2022.

[58] Zhiyuan You, Kai Yang, Wenhan Luo, Lei Cui, Yu Zheng, and Xinyi Le. Adtr: Anomaly detection transformer with feature reconstruction. *arXiv preprint arXiv:2209.01816*, 2022.

[59] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021.

[60] 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.

[61] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021.

[62] Zheng Zhang and Xiaogang Deng. Anomaly detection using improved deep svdd model with data structure preservation. *Pattern Recognition Letters*, 148:1–6, 2021.

[63] Zhenxin Zhu, Yuantao Chen, Zirui Wu, Chao Hou, Yongliang Shi, Chuxuan Li, Pengfei Li, Hao Zhao, and Guyue Zhou. Latitude: Robotic global localization with truncated dynamic low-pass filter in city-scale nerf. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8326–8332. IEEE, 2023.

[64] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12786–12796, 2022.

[65] 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.