Article

Unsupervised pose-agnostic visual anomaly detection in realistic industrial scenes

Integrated Computer-Aided Engineering
1–14

DOI: 10.1177/10692509261425164
journals.sagepub.com/home/ico

Sage | IOS Press

Enrico Marchi[1], Matic Fučka[2], Danijel Skočaj[2] and Gian Luca Foresti[1]

Abstract

Recent works on pose-agnostic anomaly detection (PAD) have addressed the challenge of identifying visual defects when the test object's pose is unknown, that is, when test images may depict the same object but in arbitrary orientations not seen in the reference anomaly-free dataset. In this unsupervised setting, models rely only on the knowledge of non-defective samples and their task is to detect anomalies appearing anywhere on the object surface. Current state-of-the-art approaches, such as OmniPoseAD, SplatPose, and SplatPose+, have advanced the field by introducing dedicated algorithms and frameworks for pose-agnostic anomaly detection. The present work consists of an engineering-oriented integration effort aimed at adapting existing PAD approaches to realistic industrial scenarios in which background clutter must be addressed for practical deployment. Two main contributions are provided: first, a simulated dataset for pose-agnostic anomaly detection with realistic industrial scenes; second, a complete pipeline that handles the introduced scenarios. Experimental results, carried out in comparison with the state-of-the-art SplatPose+ and measured in terms of pixel-level AUROC, AUPRO, image-level AUROC, and F_1-score, demonstrate good performance on the proposed dataset. Code is available at: https://github.com/enmarchi/3dpad_background.

Keywords

unsupervised anomaly detection, 3D inspection, pose-agnostic detection, gaussian

Received: 6 November 2025; accepted: 28 January 2026

1 Introduction

Modern manufacturing processes are progressively incorporating automated inspection techniques as part of the shift toward intelligent and data-driven production, with the goal of reducing human intervention in repetitive quality inspection tasks that require significant time and manpower and are susceptible to human error. In this context, research efforts are increasingly focused on developing innovative approaches for visual anomaly detection, aiming to identify structural defects through deep learning and machine vision techniques. Accordingly, this work positions itself within a broader strand of research[1–7] that integrates modern machine learning and advanced computational methods into real-world engineering systems, enabling effective and scalable solutions for practical scenarios. More specifically, the present study focuses on an in-line inspection scenario, where products move along a conveyor during the manufacturing process and are automatically analyzed to detect visual defects, determining whether each item meets quality standards or should be discarded.

When dealing with anomaly detection, defective samples are rare, which makes their collection particularly challenging. Moreover, labeling such data is a time-consuming and costly process, often requiring expert supervision. For this reason, several recent studies, such as work[8] have explored strategies to minimize the labeling effort by adopting semi-supervised or unsupervised learning approaches for visual anomaly detection. Beyond the challenges related to data scarcity and labeling, real industrial environments introduce additional sources of complexity that must be

[1]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy
[2]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

Corresponding author:
Danijel Skočaj, Faculty of Computer and Information Science, University of Ljubljana, Vecna pot 113, 1000 Ljubljana, Slovenia.
Email: danijel.skocaj@fri.uni-lj.si

carefully considered. In practice, datasets are often imperfect or contaminated with noise, such as variations in ambient lighting or reflections that affect image quality. Furthermore, despite being constrained on a conveyor system, objects may appear with slight pose mis-alignments or rotations, increasing the difficulty of consistent defect detection. Finally, complex and textured backgrounds can introduce visual artifacts that lead to false positives during object segmentation and anomaly identification, an issue that previous studies such as Stauffer and Grimson,[9] García-González[10] have attempted to address through background estimation methods.

The first contribution of this work is a publicly available dataset that simulates, through a synthetic rendering process, an industrial in-line inspection scenario, with objects moving along a conveyor and presenting varied, unpredictable poses, resulting in images with a realistic background that closely resembles an actual production environment. Regarding illumination, a uniform lighting condition was assumed, as the work does not focus on illumination-agnostic detection; nevertheless, the proposed dataset is more challenging than previous ones because it includes background clutter. Then, the aim is to identify surface anomalies that may occur at arbitrary locations on objects with unknown pose. The training dataset is assumed to contain only non-defective samples, while anomalies are artificially introduced as small and subtle defects, to evaluate the robustness of the proposed methodology under realistic and complex conditions.

The second contribution is an anomaly detection pipeline that leverages existing anomaly detection methods based on Gaussian splatting[11] and is engineered to operate on the proposed dataset.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the dataset and the proposed methodology. Section 4 presents the experimental results, followed by an ablation study. Finally, Section 5 outlines the conclusions of the paper.

2 Related works

Recent years have witnessed a growing interest in unsupervised 3D anomaly detection, driving the creation of increasingly challenging benchmarks. Early 3D anomaly detection datasets, such as MVTec 3D-AD,[12] Eyecandies,[13] and PD-Real,[14] primarily relied on single-view RGB-D data, providing only partial object observations. Subsequent datasets moved toward complete object representations: Real3D-AD[15] provides full point cloud geometry without blind spots, while 3D-ADAM[16] and Anomaly-ShapeNet[17] further enrich such representations by incorporating RGB information or synthetic reconstructions, respectively. Other dataset like MAD,[18] Real-IAD,[19] and RAD[20] achieve full object coverage through multi-view RGB images, with the latter two addressing more challenging real-world conditions. Notably, MAD was designed for pose-agnostic anomaly detection (PAD), where the object pose is unknown at test time and must be estimated before anomaly detection. PIAD[21] further extends this setup by introducing illumination changes alongside unknown poses, significantly increasing task difficulty. Furthermore, the PCAD dataset[22] bridges the gap between synthetic and real-world benchmarks by combining real multi-view RGB images with corresponding CAD models, enabling more realistic scenarios. The dataset proposed in this work shares with MAD, Real-IAD, RAD, PIAD, and PCAD the use of a multi-view approach, and furthermore consists of rendered images of synthetic LEGO models combined with synthetic yet visually realistic backgrounds.

In parallel, numerous unsupervised approaches have been proposed. Since anomalies are absent from the training set, these methods learn a model of normality solely from normal samples, classifying any deviation at inference time as anomalous. A prevalent strategy is reconstruction-based detection, where models are trained to reconstruct normal inputs but fail to accurately reproduce anomalies, as in works.[17,23–27] In addition, several works[24,28–31] attempt to leverage RGB information by fuzing it with depth data. In other approaches,[32–37] the architecture generates simulated anomalies on normal images for training purposes; it then learns to reconstruct the normal appearance from the anomalous image, and at inference time employs the extracted features to drive the anomaly detection process using the reconstructed normal object as a guide. Other studies[38–43] adopt a different approach based on the paradigm introduced by PatchCore,[44] in which features from normal images are extracted during training and stored in a feature dictionary, referred to as a memory bank. At test time, features extracted from the query item are compared against the memory bank, and an anomaly score is assigned based on their distance from the stored features.

Gaussian splatting[11] was introduced as a method for 3D scene modeling using RGB images with known camera poses, which enables rendering images from arbitrary viewpoints. Since its introduction, it has been applied to the PAD problem. The first approaches, SplatPose[45] and IGSPAD,[46] leverage the differentiability of Gaussian splatting to design a pose estimation module that solves an optimization problem with respect to the pose, a strategy that is also adopted in PIAD.[21] In contrast, SplatPose+[47] departs from this idea and estimates object poses at test time via PnP[48]-based optimization. The latter is then adapted in the proposed pipeline to handle background content, which is not considered in previous related approaches.

3 The proposed system

The proposed approach builds upon the SplatPose+ framework, which is briefly reviewed in the next subsection. The

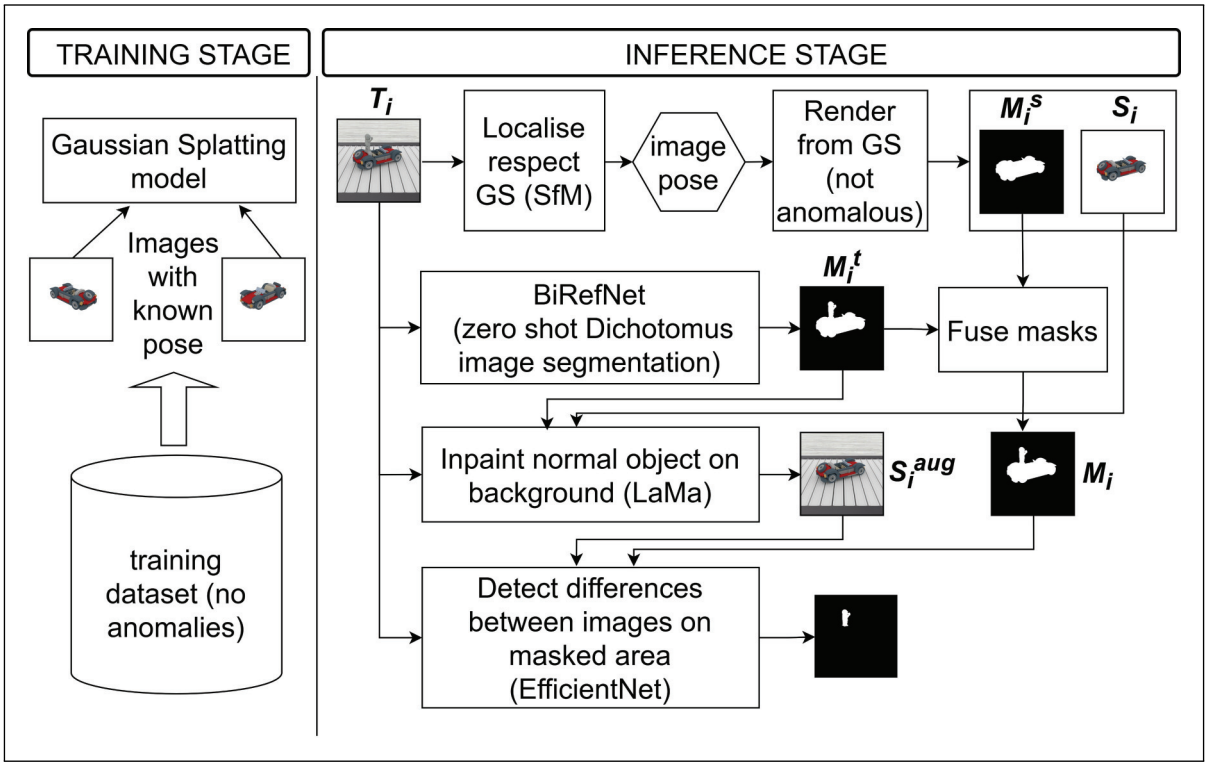

Figure 1. Proposed architecture.

subsequent subsections detail the proposed dataset and the original pipeline.

3.1 Pose estimation and anomaly detection

In SplatPose+, the 3D representation of non-anomalous models is learned through Gaussian splatting, where the scene is modeled as a collection of splats, characterized by parameters such as 3D position, color, opacity and 3D shape.

The training process consists of optimizing the parameters of these splats so that the rendered views of the reconstructed model closely match the input RGB images.

To initialize the Gaussian splatting model, a preliminary 3D reconstruction is obtained using Structure-from-Motion (SfM).[49,50] The resulting SfM point cloud is scale-consistent and therefore provides an approximate yet reliable representation of the object geometry; each 3D point, containing only positional information, is used as the initial location of a Gaussian splat. Subsequent optimization refines these splats, enriching them with additional attributes such as color, opacity, and spatial extent.

In addition, SplatPose+ leverages the optimized SfM reconstruction to estimate test-time camera poses via PnP optimization and anomaly detection is then performed directly in the image domain. Specifically, a synthetic view of the object is rendered from the same viewpoint using the learned Gaussian splatting model, which represents the normal, defect-free appearance. This rendered image serves as a reference and is compared with the corresponding real test image. Both images are processed through a feature extraction backbone, and the resulting feature maps are analyzed to measure their discrepancy. Regions showing significant differences between the two sets of features indicate deviations from normality, allowing the detection and localization of surface anomalies.

Algorithm 1. Overview of the proposed pipeline.

Require: Test image T_i, synthesized view S_i

$M_i^t \leftarrow \text{BiRefNet}(T_i)$

$M_i^s \leftarrow \text{ExtractReferenceMask}(S_i)$

$M_i \leftarrow M_i^t \vee M_i^s$

$S_i^{\text{aug}} \leftarrow \text{Compose}(S_i|_{M_i^s}, \text{LaMa}(T_i, M_i^t))$

$\text{AnomalyDetection}(T_i|_{M_i}, S_i^{\text{aug}}|_{M_i})$

3.2 Methodology

The proposed pipeline is shown in Figure 1 and described in Algorithm 1. In this work, particular attention is devoted to addressing the pose-agnostic anomaly detection problem under conditions that more closely resemble real industrial environments. The main challenge arises from background elements in the test images. Because objects have no fixed pose, neither their pose is known a-priori nor is their arrangement with respect to the background

consistent across different objects. As a result, training a Gaussian splatting normal model that includes background content and comparing it to a test image with background is unreliable: even when the object view matches, the background pose does not, and vice versa. To mitigate this, the processing pipeline aims to remove the influence of the background. Accordingly, the Gaussian splatting model is trained on images with a null background, represented by a uniform white surface, as in previous published works. To ensure a consistent comparison between rendered and test images, the test background must be suppressed. This requires an accurate object segmentation and restricting the comparison to the object mask, thereby eliminating background influence.

As will be explained in the dataset section, anomalies can involve missing, modified, or added parts. In the case of missing parts, detection must focus on the region where the part should be. The object mask from the test image is therefore insufficient, because it can exclude the removed part, as would happen in the example shown in Figure 2(a) where the removed part is above the background. Then, it is necessary to also use the object mask from the synthesized normal image, which delineates the expected extent of the part, so that the corresponding area can be analyzed.

On the contrary, for added parts, relying only on the synthesized normal mask would miss the anomalous region, since anomalies can lie outside the expected support, like in the example shown in Figure 2(b); the test-image object mask must therefore be used to include those extra areas. In practice, the union of the two masks is analyzed: regions present only in the normal mask indicate missing parts, regions present only in the test mask indicate added parts, and discrepancies within the overlap correspond to modified parts.

Formally, let M_i^s denote the object mask in the synthesized view S_i and M_i^t the object mask in the test image T_i. In the proposed example Figure 3(a) corresponds to T_i, while Figure 3(b) corresponds to S_i. The mask M_i^s is obtained by thresholding the final residual transmittance of the synthesized image. Indeed, in Gaussian splatting, the color of pixel p_j is obtained by front-to-back alpha compositing of the splats' colors c_k^j weighted by their opacities $\alpha_k^j \in [0, 1]$ and the residual transmittance $T_k^j \in [0, 1]$. The opacity α_k^j is given by the value of the 2D Gaussian associated with the kth splat at pixel j, multiplied by its learned opacity:

$$p_j = \sum_{k=1}^{N_j} T_k^j \, \alpha_k^j \, c_k^j, \tag{1}$$

where N_j is the number of splats contributing to pixel j, and

$$T_k^j = \prod_{h<k} \left(1 - \alpha_h^j\right), \tag{2}$$

with splats ordered by increasing depth.

Intuitively, Tr_k^j measures the residual mass of the pixel j available to be filled after the accumulation of the first $k-1$ contributions; when the rendering process is terminated $Tr_N \approx 1$ on background pixels, while on pixels representing the object $Tr_N \ll 1$. Therefore, a clean foreground–background separation for the synthesized image is obtained by thresholding Tr.

To obtain M_i^t, a BiRefNet[51] architecture is used. BiRefNet is a zero-shot model for dichotomous image segmentation, designed to separate foreground from background. Then M_i^s and M_i^t are fuzed by a pixel-wise logical OR to obtain the final mask $M_i = M_i^s \vee M_i^t$, which specifies the region over which anomaly detection is performed when comparing the test and synthesized images. In the proposed example, Figure 3(c) corresponds to M_i.

This is still not sufficient, as residual errors depend on mask accuracy. In particular, if the mask spills beyond the foreground object, the comparison will include background regions, which may be detected as false anomalies. To mitigate this issue, an inpainting strategy is applied to suppress background content within the comparison region. Specifically, the test image T_i is processed using the LaMa model,[52] which removes the object and reconstructs the underlying background. The inpainting region is defined by the mask M_i^t. Once background reconstruction is complete, the anomaly-free object from the synthesized image S_i is extracted using M_i^s and composited onto the reconstructed background, yielding S_i^{aug}. In the proposed example Figure 3(d) corresponds to S_i^{aug}.

Finally, the anomaly detection is performed between T_i and S_i^{aug} restricted to the region defined by M_i. Figure 3(e) shows the detected anomaly in the proposed example.

3.3 Implementation

As described in SplatPose+, an SfM model is built from the training images to initialize Gaussian splatting and to localize test images, using NetVLAD[53] for retrieval, SuperPoint[54] and LightGlue[55] for matching, and hloc[50] for triangulation. The Gaussian splatting model is subsequently trained under the same settings, with 15000 iterations, a densification interval of 1000 iterations, and a spherical harmonics degree of 3.

Anomaly detection is performed as in OmniposeAD,[18] where features are extracted using a pre-trained backbone (EfficientNet[56]) on both the synthesized reference and the test image. In this work, however, the comparison is carried out using the normalized L_1 distance between paired features at each of the five layers, which is less sensitive to noise than the L_2 metric adopted in OmniposeAD. All resulting maps are resized to 224×224, summed across layers, Gaussian-smoothed, and normalized to obtain the pixel-wise anomaly map. The image-level anomaly score is the maximum value of this map.

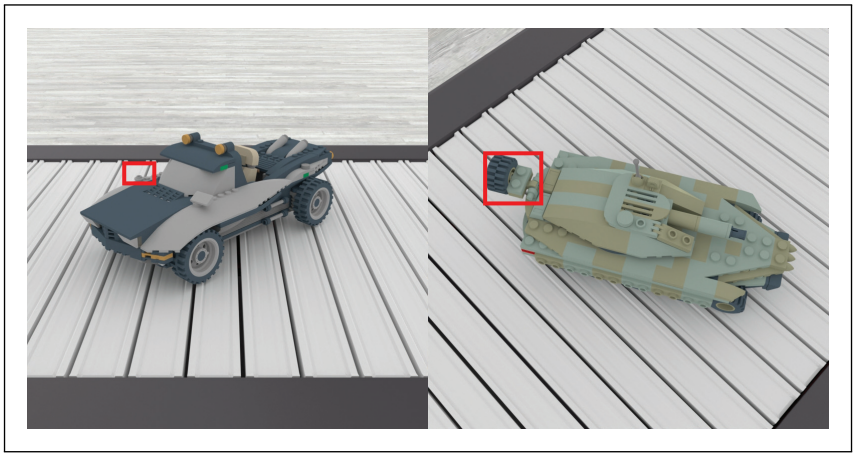

Figure 2. Left: (a) Removed part (mirror) visible against the background (red box). Right: (b) Added defect visible against the background (red box).

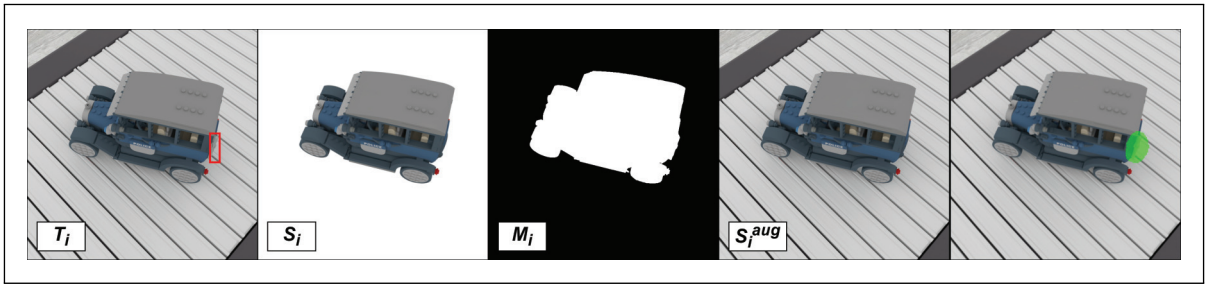

Figure 3. From left to right: (a) test image with the anomaly highlighted (box); (b) synthetic normal image; (c) foreground mask; (d) inpainted normal object on the reconstructed background; (e) anomaly detection image.

For background reconstruction with LaMa, the inpainting region is obtained by dilating the mask M_i^t by 5 pixel. Indeed, dilating the mask avoids operating exactly on the object boundary, where recovering the underlying background is more difficult. After reconstruction, anomaly detection compares T_i with S_i^{aug}. In principle, this comparison could be performed over the entire image, since the anomaly-free object is composited onto the reconstructed background. However, because the inpainted region is dilated and the reconstruction, while generally accurate, may still contain local imperfections, evaluating everywhere could introduce false anomalies due to inpainting artifacts. To minimize such effects, the comparison is restricted to the minimal necessary region, namely the mask M_i.

3.4 Dataset

The dataset used in this work was developed in Blender,[57] which enables the creation of consistent industrial scenarios.

Following the approach introduced in the MAD dataset, the proposed dataset was built using LEGO-based models specifically designed by expert modelers, who provided the original, defect-free base structures. Nine distinct vehicle models were employed which are police car, coupe car, amphibious car, off-road car, tank, desert car, truck, loader and racing car; these models were chosen to represent objects with increasing structural complexity. While simpler models such as the coupe exhibits compact and mostly closed shapes, the amphibious car, the tank, the truck, the loader, the off-road car, the racing car, the desert car introduce more intricate geometries, and the police car model features a high number of detailed parts and open regions. This progression from simple to complex shapes not only increases the number of components but also introduces concave and perforated structures, where background visibility through small openings may occur, potentially leading to false positives in anomaly detection.

For each class, corresponding to one vehicle model, 100 defect-free training images were rendered with known intrinsic and extrinsic camera parameters defining the pose

Table 1. Percentage of anomalous test images per class.

	Anomalous (%)
Amphibious car	44
Coupe car	45
Police car	41
Tank car	46
Off-road car	42
Desert car	39
Loader car	36
Racing car	41
Truck car	36

P. The camera viewpoints were distributed over the visible upper hemisphere of the object which is the part not in contact with the conveyor.

Additionally, 13 further objects per class were generated for testing and each of them was placed in a random pose distinct from the pose of the training object. Seven of them were modified by adding small, subtle surface defects to simulate realistic flaws. The considered defects fall into the categories of additions, removals, and color alterations, which are implemented by respectively adding, deleting, or replacing LEGO components. Specifically, for each class, two of the modified objects exhibit color changes, two present removed components, and three include added components. The remaining six objects were left intact, representing non-defective items with unseen orientations; this setup was intended to simulate a realistic production scenario in which normal items are also inspected and should not be classified as anomalous. For each test object, 9 RGB images at 800×800 resolution were captured with unknown poses, and therefore a total of 117 test images were obtained for each class. Table 1 reports, for each class, the percentage of anomalous test images.

The modifications were intentionally kept minimal with respect to the overall object volume, so that the proposed method can be evaluated under challenging conditions. An example is shown in Figure 4(a). Moreover, some anomalies were made larger, especially in the case of added parts. This choice was made to account for perspective effects: depending on the viewpoint, anomalies may appear either over the object or over the background as in Figure 4(b). By introducing sufficiently large anomalies, it becomes more likely to include cases where the added parts appear on the background rather than always overlapping the object itself. Indeed, if anomalies overlap just the object and not the background, their detection is simplified, as background clutter is eliminated.

During rendering, the test images included a realistic background, whereas the training images were rendered on a white background. This choice allowed the construction of reference normal models representing only the object of interest, without the background, which would be unnecessary and possibly misleading for the anomaly detection stage. However, to maintain consistent illumination between the training and test data, all renders were produced within the same scene and lighting setup; the training objects were subsequently cropped and placed on a white background, which is a reasonable assumption since training data can be pre-processed offline. The background scene consisted of a conveyor model and a textured floor, designed to avoid monochromatic regions while keeping the setup minimal. The conveyor was modeled as a series of metallic plates, providing a moderately complex yet regular texture. Moreover, for each test object, the conveyor position was kept fixed across its rendered views but slightly shifted between different objects, ensuring that the background pattern remained consistent within the same object while varying across samples. This promotes generalization and reduces background bias.

4 Experimental results

The following sections present quantitative and qualitative evaluations comparing the proposed method with the state-of-the-art SplatPose+.

4.1 Metrics

In accordance with prior works,[18,45,47] the evaluation metrics includes pixel-level AUROC, AUPRO,[12] and image-level AUROC, where image-level anomaly score is defined as the maximum of the pixel-wise anomaly map. Higher values of pixel-level AUROC and AUPRO indicate better localization of true anomalies, as explicitly assess the spatial accuracy of the pixel-wise map.

While the previous metrics are threshold-free, qualitative visualization and practical anomaly detection require setting a decision threshold, which in turn enables the computation of the F_1-score to measure the quality of anomaly detection at localization level. The proposed pipeline produces a soft anomaly mask at the pixel level and aggregates it into a single image-level score by taking the maximum value across pixels. Then, a decision threshold is selected by maximizing the image-level F_1-score. Therefore, an image is considered anomalous if its aggregated score

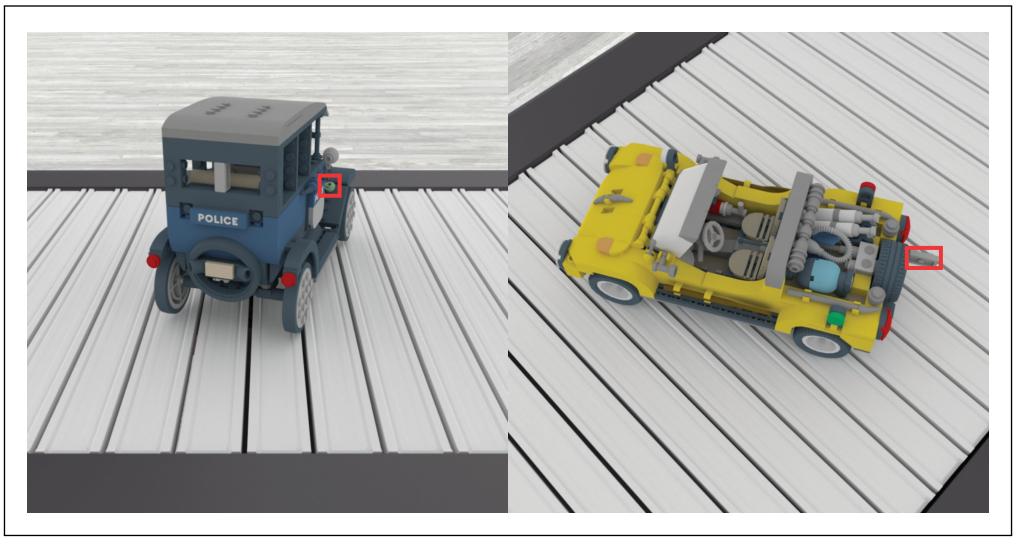

Figure 4. Left: (a) Small added defect (box). Right: (b) Added defect visible against the background (box).

exceeds this threshold; the same threshold is then applied at pixel level, classifying pixels with scores above it as anomalous. With this in mind it is possible to compute the F_1-score at the anomaly-localization level as follows. After thresholding the pixel-wise anomaly map, connected components are extracted and the same procedure is applied to the ground-truth mask. A true positive area is counted when a predicted connected component overlaps with any ground-truth one. A false positive area is a predicted connected component that does not overlap with any ground-truth one. A false negative area is a ground-truth connected component that does not overlap with any predicted one. Since each image contains at most one anomaly, all its ground-truth connected components are treated as a single anomaly instance. Therefore, if at least one ground-truth component is intersected by any prediction, the anomaly is counted as a single true positive and no false negative is recorded for that image, regardless of how many components remain uncovered. Likewise, multiple predicted components overlapping the same anomaly are still counted as a single true positive. F_1-score is then computed using the number of true positives, false positives and false negatives. The F_1-score also measures how accurate the image-level classification is; specifically, whether an image labeled as anomalous reflects the detection of a genuine anomaly or merely spurious, noise-induced false positives.

4.2 Quantitative evaluation

Table 2 reports the performance of the proposed pipeline and the per-image inference time for all nine classes, based on experiments run on an NVIDIA RTX-A4500 GPU. The results are reported up to the third decimal digit and this level of precision is required especially for pixel-level AUROC and AUPRO because the anomalous pixel class is unbalanced relative to the non-anomalous class. High numerical precision is therefore required to capture small variations in the detection of fine anomalies.

To the best of the authors' knowledge, all existing methods addressing the PAD problem assume that backgrounds are removed from test images. A direct comparison with such approaches would therefore be misleading, as their performance is expected to degrade substantially under the proposed setting, which includes background clutter. Then, to obtain a meaningful point of reference to the state-of-the-art SplatPose+, the comparison is conducted by evaluating SplatPose+ on a simplified dataset with same objects and poses, but with backgrounds removed, whereas the proposed pipeline is evaluated on the original dataset that retains background clutter. The results show that the proposed pipeline attains comparable performance to SplatPose+ in this setting, indicating that any remaining gap is not due to the background-handling challenge addressed by the pipeline, but reflects difficulty intrinsic to the task even under simplified conditions. The resulting scores, are written in Table 3. Table 4 shows, instead, the per-class F_1-scores for both the proposed method and SplatPose+. Results are reported to two decimal places, which is, in this case, sufficiently informative.

All evaluations are averaged over multiple runs of the same pipeline and results are reported as mean and standard deviation. The variability is intrinsic to the Gaussian splatting technique. During model optimization, new splats may be introduced with a small degree of randomness, which can lead to slight differences across runs even when training on the same data.

Overall, the performances between methods are comparable even though the proposed method is evaluated on original images with background clutter, unlike SplatPose+, which is tested on background-removed images.

Table 2. Proposed method evaluation results and inference times.

	p-AUROC	AUPRO	i-AUROC	i-time (s)
Amphibious car	0.999 ± 0.000	0.994 ± 0.000	0.985 ± 0.002	1.7 ± 0.0
Coupe car	0.999 ± 0.000	0.993 ± 0.000	0.986 ± 0.001	1.7 ± 0.0
Police car	0.992 ± 0.000	0.985 ± 0.000	0.921 ± 0.006	1.7 ± 0.1
Tank car	0.999 ± 0.000	0.990 ± 0.000	0.987 ± 0.001	1.7 ± 0.0
Off-road car	0.999 ± 0.000	0.994 ± 0.000	0.981 ± 0.002	1.7 ± 0.0
Desert car	0.999 ± 0.000	0.995 ± 0.000	0.940 ± 0.003	1.7 ± 0.0
Loader car	0.999 ± 0.000	0.997 ± 0.000	0.993 ± 0.003	1.7 ± 0.0
Racing car	0.996 ± 0.000	0.977 ± 0.000	0.947 ± 0.005	1.7 ± 0.0
Truck car	0.999 ± 0.000	0.993 ± 0.000	0.972 ± 0.003	1.7 ± 0.0

Note: Anomaly detection metrics: pixel-wise AUROC (p-AUROC), AUPRO, image-wise AUROC (i-AUROC) and image inference time (i-time) in seconds.

Table 3. SplatPose+[a] results.

	p-AUROC	AUPRO	i-AUROC	i-time (s)
Amphibious car	0.999 ± 0.000	0.994 ± 0.000	0.989 ± 0.002	0.9 ± 0.0
Coupe car	0.999 ± 0.000	0.991 ± 0.000	0.992 ± 0.001	0.9 ± 0.0
Police car	0.999 ± 0.000	0.993 ± 0.000	0.921 ± 0.011	0.9 ± 0.0
Tank car	0.998 ± 0.000	0.984 ± 0.000	0.977 ± 0.002	0.9 ± 0.0
Off-road car	0.999 ± 0.000	0.993 ± 0.000	0.986 ± 0.001	0.9 ± 0.0
Desert car	0.999 ± 0.000	0.994 ± 0.000	0.961 ± 0.001	0.9 ± 0.0
Loader car	0.999 ± 0.000	0.996 ± 0.000	0.993 ± 0.004	0.9 ± 0.0
Racing car	0.998 ± 0.000	0.989 ± 0.000	0.956 ± 0.003	0.9 ± 0.0
Truck car	0.999 ± 0.000	0.994 ± 0.000	0.981 ± 0.002	0.9 ± 0.0

Note: Anomaly detection metrics: pixel-wise AUROC (p-AUROC), AUPRO, image-wise AUROC (i-AUROC) and image inference time (i-time) in seconds.
[a]Evaluated on images with the background removed.

Table 4. Per-class F_1-scores for the proposed method and SplatPose+[a].

	Proposed method	SplatPose+
Amphibious car	0.92 ± 0.01	0.95 ± 0.01
Coupe car	0.94 ± 0.00	0.98 ± 0.00
Police car	0.79 ± 0.02	0.78 ± 0.04
Tank car	0.92 ± 0.01	0.93 ± 0.01
Off-road car	0.92 ± 0.01	0.90 ± 0.02
Desert car	0.83 ± 0.02	0.83 ± 0.02
Loader car	0.95 ± 0.02	0.93 ± 0.02
Racing car	0.88 ± 0.02	0.89 ± 0.01
Truck car	0.88 ± 0.02	0.94 ± 0.01

[a]Evaluated on images with the background removed.

The easiest class, coupe car, achieves the best results, as expected, since it is relatively compact and lacks small holes or irregularities in its shape that could degrade the anomaly detection performance.

Conversely, complex objects, such as the police car present fine structural details like small open regions that make accurate detection more challenging, lowering the scores. More specifically, concave regions and small holes may be incorrectly detected by BiRefNet as part of the object rather than as background. In such cases, LaMa is employed to mitigate this issue; however, when holes are located within the object and are very small, the inpainting technique may still be insufficient to accurately reconstruct the background in these regions, as shown in Figure 6. These aspects represent the main limitations of the proposed pipeline. However, SplatPose+ also exhibits limitations when dealing with small holes or concavities, as illustrated in Figure 6. In particular, the Gaussian splatting technique may fail to accurately reconstruct small holes and cavities during training; as a consequence, during anomaly detection, these regions can be incorrectly identified as anomalous, leading to false positives.

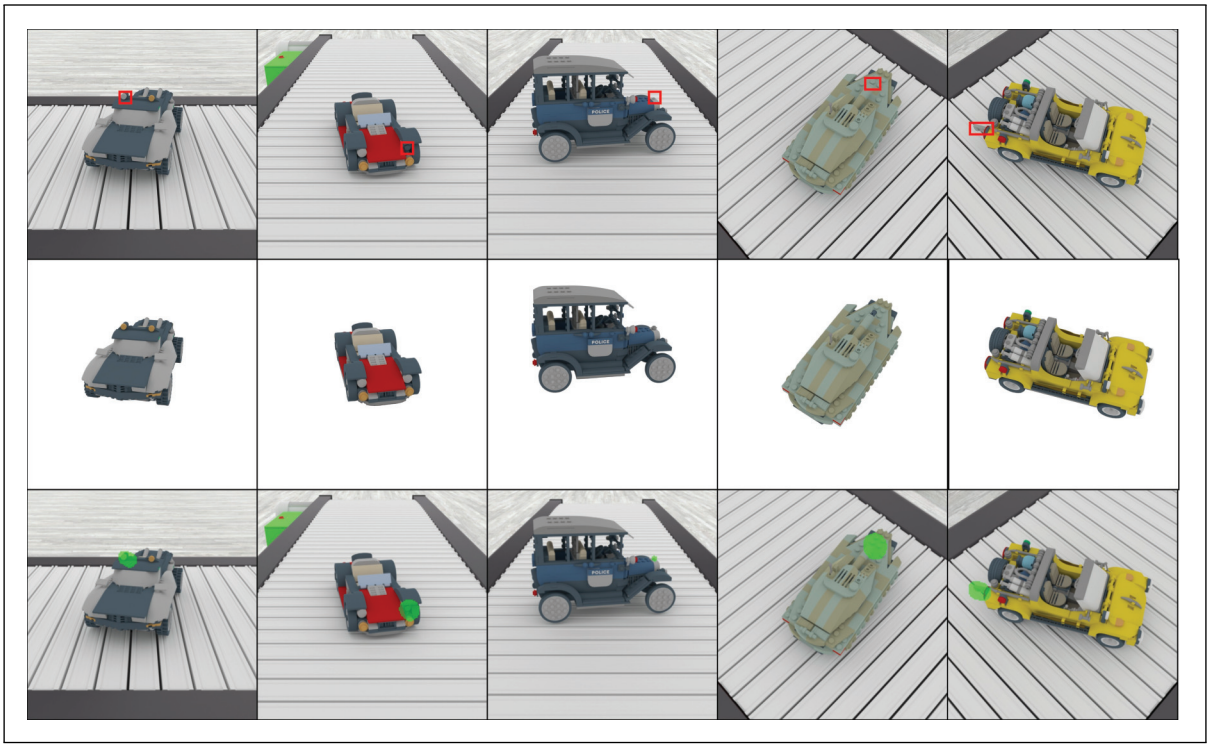

Figure 5. Examples for each class. Top row: testing images. Middle row: synthetic normal images. Bottom row: image with detected anomaly.

The results may appear sometimes slightly inconsistent, since for some classes a higher image-level AUROC does not necessarily correspond to a higher F_1-score. This is due to the fact that image-level AUROC disregards anomaly localization and can be affected by noise.

Regarding computational time, the background subtraction stage implemented using the LaMa and BiRefNet modules introduces an average inference time of approximately 0.8 seconds per class. This is consistent with the results reported in Tables 2 and 3, where the last column shows a difference of 0.8 seconds between the proposed pipeline and SplatPose+.

4.3 Qualitative evaluation

The masked regions in the bottom row of Figure 5 indicate anomalous pixels obtained by applying the previously defined decision threshold, while the first and second rows show the test images and the synthesized normal images, respectively. Figure 6 highlights, within boxes, examples of false positives and false negatives for the proposed method in the first row and for SplatPose+ in the second row; the third row displays the ground-truth masks. The images show that both approaches can produce spurious or missing detections.

4.4 Ablation study

Finally, two ablation studies are conducted. The first analyzes the impact of the LaMa inpainting and BiRefNet segmentation modules. The second reports the results obtained by running the proposed pipeline on background-free images, following the SplatPose+ setting.

Table 6 reports results obtained when comparing test and reference images only within the object mask of the synthesized normal image, which is the mask derived from residual transmittance. Performance remains strong because, in many views, anomalies are small and lie well inside the object rather than near its boundaries, making this mask sufficient in numerous cases. However, this cannot always be assumed: in worst-case scenarios where anomalies occur on the border, restricting the comparison to the synthesized mask misses relevant regions. In such cases, the full proposed pipeline is required to recover the complete comparison area. Consistently, Table 2 shows improved results, as it also addresses the border cases discussed above. Table 5 reports the F_1-score performance and should be compared with the first column of Table 4. For classes such as amphibious car and coupe car, the results are essentially comparable across methods. For the remaining classes, however, the proposed method attains higher F_1-scores, highlighting the improvements delivered by the proposed approach.

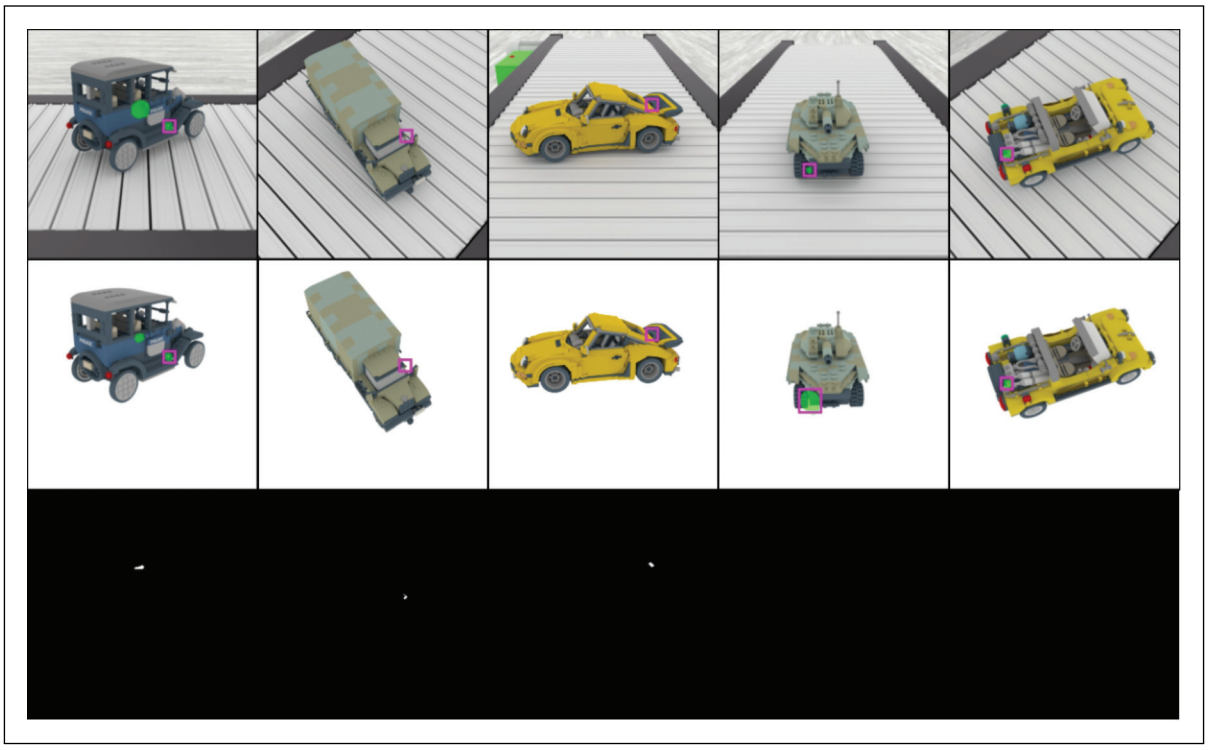

Figure 6. Qualitative comparison of detection errors. Examples of false positives and false negatives are highlighted with boxes. Top row: proposed method; middle row: SplatPose+; bottom row: ground-truth masks.

Table 5. Per-class F_1-scores obtained without employing BiRefNet and LaMa modules.

	F_1
Amphibious car	0.91 ± 0.01
Coupe car	0.94 ± 0.00
Police car	0.70 ± 0.04
Tank car	0.89 ± 0.01
Off-road car	0.85 ± 0.01
Desert car	0.81 ± 0.01
Loader car	0.87 ± 0.02
Racing car	0.79 ± 0.05
Truck car	0.84 ± 0.02

Table 6. Results and inference times obtained without employing BiRefNet and LaMa modules.

	p-AUROC	AUPRO	i-AUROC	i-time (s)
Amphibious car	0.993 ± 0.000	0.983 ± 0.001	0.975 ± 0.004	0.9 ± 0.0
Coupe car	0.983 ± 0.000	0.980 ± 0.000	0.988 ± 0.001	0.9 ± 0.0
Police car	0.977 ± 0.000	0.940 ± 0.002	0.873 ± 0.009	0.9 ± 0.0
Tank car	0.956 ± 0.000	0.956 ± 0.000	0.974 ± 0.002	0.9 ± 0.0
Off-road car	0.971 ± 0.000	0.916 ± 0.001	0.911 ± 0.006	0.9 ± 0.0
Desert car	0.989 ± 0.001	0.961 ± 0.002	0.907 ± 0.005	0.9 ± 0.0
Loader car	0.979 ± 0.001	0.938 ± 0.002	0.939 ± 0.007	0.9 ± 0.0
Racing car	0.980 ± 0.000	0.927 ± 0.001	0.899 ± 0.008	0.9 ± 0.0
Truck car	0.977 ± 0.001	0.957 ± 0.002	0.930 ± 0.007	0.9 ± 0.0

Note: Anomaly detection metrics: pixel-wise AUROC (p-AUROC), AUPRO, image-wise AUROC (i-AUROC) and image inference time (i-time) in seconds.

Table 7. Evaluation results of the proposed method on background-free images.

	p-AUROC	AUPRO	i-AUROC	F_1
Amphibious car	0.999 ± 0.000	0.993 ± 0.000	0.983 ± 0.002	0.91 ± 0.02
Coupe car	0.999 ± 0.000	0.991 ± 0.000	0.991 ± 0.001	0.92 ± 0.01
Police car	0.999 ± 0.000	0.992 ± 0.000	0.920 ± 0.008	0.79 ± 0.04
Tank car	0.999 ± 0.000	0.984 ± 0.000	0.980 ± 0.002	0.92 ± 0.02
Off-road car	0.999 ± 0.000	0.992 ± 0.000	0.987 ± 0.001	0.90 ± 0.02
Desert car	0.999 ± 0.000	0.993 ± 0.000	0.958 ± 0.002	0.86 ± 0.00
Loader car	0.999 ± 0.000	0.996 ± 0.000	0.993 ± 0.003	0.96 ± 0.01
Racing car	0.999 ± 0.000	0.990 ± 0.000	0.961 ± 0.005	0.88 ± 0.01
Truck car	0.999 ± 0.000	0.993 ± 0.000	0.979 ± 0.004	0.93 ± 0.02

Note: Anomaly detection metrics: pixel-wise AUROC (p-AUROC), AUPRO, image-wise AUROC (i-AUROC) and F_1-score.

Table 7 reports the usual performance metrics of the proposed pipeline when applied to background-free images. The results are comparable to those obtained on images with background and to the performance achieved by SplatPose+ on background-free images.

5 Conclusion

A dataset for pose-agnostic anomaly detection was introduced to reflect realistic industrial inspection conditions by incorporating a realistic, yet synthetic, cluttered background, along with a pipeline engineered by integrating existing modules such as SplatPose+, LaMa, and BiRefNet to address the practical constraints inherent in the data The proposed method is able to detect anomalies, although some aspects remain unresolved.

A primary issue concerns the image-level decision rule. Small amounts of random noise can produce false positives, because nothing prevents the noise features from exhibiting large magnitude differences relative to normality. Since the aggregation uses the maximum across pixel scores, a single spurious peak can cause a normal image to be classified as anomalous. Similar failures arise in small surface holes. These regions are difficult to separate from the object, often lie within the object silhouette, and are compared against non-anomalous exemplars that exhibit blank white backgrounds in those holes. This mismatch can introduce false positives. The proposed pipeline limits many of these cases, but performance is not yet perfect. Further investigation is required to refine these borderline scenarios and to improve robustness. Moreover, the computational time could be improved, and further research is needed to reduce it.

This work has some limitations, most notably the fact that the proposed approach is evaluated in a controlled synthetic environment. As a result, there is room for future work aimed at developing a pipeline that is closer to real industrial scenarios, where noise is more pronounced. Furthermore, illumination conditions, which were assumed to be uniform in this study, should be addressed in future work to reduce sensitivity to lighting changes and to variability arising from the temporal degradation of equipment and environmental conditions.

Acknowledgements

The authors gratefully acknowledge the support of Łukasz Leon Grzywacz (https://www.seymouria.pl/) for providing the 3D LEGO-based models used in the dataset. ChatGPT was used exclusively throughout the text for grammatical review (e.g., punctuation, singular/plural agreement, and verb tenses consistency) and to assess sentence clarity; it was not used to generate any original content. This paper was partially supported by University of Udine project on "Piano Stretegico Dipartimentale on Artificial Intelligence" (PSD-AI) (2022-25) project at the University of Udine.

Ethical approval and informed consent

Not applicable. This study did not involve human participants, human data, or animal experiments.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability

Dataset is available at: https://figshare.com/s/6584055d09cad97c949c.

ORCID iDs

Enrico Marchi https://orcid.org/0009-0004-6090-7891
Matic Fučka https://orcid.org/0009-0007-9654-8386

Gian Luca Foresti https://orcid.org/0000-0002-8425-6892
Danijel Skočaj https://orcid.org/0000-0002-5290-4736

References

1. Newman TS and Jain AK. A survey of automated visual inspection. *Comput Vis Image Underst* 1995; 61: 231–262.
2. Garrido-Hidalgo C, Roda-Sanchez L, Fernández-Caballero A, et al. Internet-of-things framework for scalable end-of-life condition monitoring in remanufacturing. *Integr Comput Aided Eng* 2024; 31: 1–17.
3. Michailidis P, Michailidis IT, Gkelios S, et al. Neuro-distributed cognitive adaptive optimization for training neural networks in a parallel and asynchronous manner. *Integr Comput Aided Eng* 2024; 31: 19–41.
4. Neri F, Liu X, Yue P, et al. Deep deterministic policy gradient with constraints for gait optimisation of biped robots. *Integr Comput Aided Eng* 2024; 31: 139–156.
5. Zhou C, Fan L and Neri F. A spatio-temporal fusion deep learning network with application to lightning nowcasting. *Integr Comput Aided Eng* 2024; 31: 233–247.
6. Marcondes FS, Almeida JJ and Novais P. An exploratory design science research on troll factories. *Integr Comput Aided Eng* 2024; 31: 95–115.
7. Marchi E, Fornasier D, Miorin A, et al. Segmentation networks for detecting overlapping screws in 3D and color images for industrial quality control. *Integr Comput Aided Eng* 2025; 32: 244–257.
8. Krassnig PJ, Haselmann M, Kremnitzer M, et al. Efficient surface defect detection in industrial screen printing with minimized labeling effort. *Integr Comput Aided Eng* 2024; 32: 3–23.
9. Stauffer C and Grimson WEL. Adaptive background mixture models for real-time tracking. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Volume 2, pp.246–252. IEEE. https://doi.org/10.1109/CVPR.1999.784637.
10. García-González J, Ortiz-de Lazcano-Lobato JM, Luque-Baena RM, et al. Background subtraction by probabilistic modeling of patch features learned by deep autoencoders. *Integr Comput Aided Eng* 2020; 27: 253–265.
11. Kerbl B, Kopanas G, Leimkuehler T, et al. 3D gaussian splatting for real-time radiance field rendering. *ACM Trans Graph* 2023; 42: 1–14.
12. Bergmann P, Jin X, Sattlegger D, et al. 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)*. https://doi.org/10.5220/0010865000003124.
13. Bonfiglioli L, Toschi M, Silvestri D, et al. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In: Wang L, Gall J, Chin TJ, Sato I and Chellappa R (eds) *Computer vision – ACCV 2022*. Cham: Springer Nature Switzerland, pp.459–475. https://doi.org/10.1007/978-3-031-26348-4_27.
14. Qin J, Gu C, Yu J, et al. Image-pointcloud fusion based anomaly detection using PD-REAL dataset. arXiv preprint arXiv:231104095, 2023. https://arxiv.org/pdf/2311.04095.pdf.
15. Liu J, Xie G, Chen R, et al. Real3D-AD: a dataset of point cloud anomaly detection. In: *Proceedings of the 37th international conference on neural information processing systems (NeurIPS 2023)*. NeurIPS '23. Red Hook, NY, USA: Curran Associates, Inc. https://doi.org/10.5555/3666122.3667446.
16. McHard P, Audonnet FP, Summerell O, et al. 3D-ADAM: a dataset for 3D anomaly detection in additive manufacturing. arXiv preprint arXiv:250707838, 2025. https://arxiv.org/pdf/2507.07838.pdf.
17. Li W, Xu X, Gu Y, et al. Towards scalable 3D anomaly detection and localization: a benchmark via 3D anomaly synthesis and a self-supervised learning network. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.22207–22216. https://doi.org/10.1109/CVPR52733.2024.02096.
18. Zhou Q, Li W, Jiang L, et al. PAD: a dataset and benchmark for pose-agnostic anomaly detection. In: *Proceedings of the 37th international conference on neural information processing systems (NeurIPS 2023)*. NeurIPS '23. Red Hook, NY, USA: Curran Associates, Inc. https://doi.org/10.5555/3666122.3668052.
19. Wang C, Zhu W, Gao B, et al. Real-IAD: a real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In: *2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.22883–22892. IEEE. https://doi.org/10.1109/CVPR52733.2024.02159.
20. Zhou K, Cao Y, Kim T, et al. RAD: a dataset and benchmark for real-life anomaly detection with robotic observations. arXiv preprint arXiv:241000713, 2024. https://arxiv.org/pdf/2410.00713.pdf.
21. Yang K, Cao J, Bai Z, et al. PIAD: pose and illumination agnostic anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.4734–4743.
22. Maack R, Thun L, Liang T, et al. PCAD: a real-world dataset for 6d pose industrial anomaly detection. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision workshops (WACVW)*, pp.1132–1141.
23. Gencer Z, Sahin YH and Unal G. One-class classification of 3D point clouds using dynamic graph CNN. In: *Proceedings of the 7th international conference on computer science and engineering (UBMK)*, pp.388–392. https://doi.org/10.1109/UBMK55850.2022.9919505.
24. Rudolph M, Wehrbein T, Rosenhahn B, et al. Asymmetric student-teacher networks for industrial anomaly detection. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV)*, pp.2591–2601. https://doi.org/10.1109/WACV56688.2023.00262.

25. Bergmann P and Sattlegger D. Anomaly detection in 3D point clouds using deep geometric descriptors. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision (WACV)*, pp.2612–2622. https://doi.org/10.1109/WACV56688.2023.00264.
26. Lhoste R, Vacavant A and Delhay D. MAESTRO: a full point cloud approach for 3D anomaly detection based on reconstruction. In: *Proceedings of VISIGRAPP 2025, Volume 2: VISAPP*, pp.717–724. https://doi.org/10.5220/0013250500003912.
27. Masuda M, Hachiuma R, Fujii R, et al. Toward unsupervised 3D point cloud anomaly detection using variational autoencoder. In: *Proceedings of the IEEE international conference on image processing (ICIP)*, pp.3118–3122. https://doi.org/10.1109/ICIP42928.2021.9506795.
28. Wang J, Niu Y and Huang B. Fusion-restoration model for industrial multimodal anomaly detection. *Neurocomputing* 2025; 637: 130073.
29. Costanzino A, Ramirez PZ, Lisanti G, et al. Multimodal industrial anomaly detection by crossmodal feature mapping. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern 4ecognition (CVPR)*. pp.17234–17243. https://doi.org/10.1109/CVPR52733.2024.01631.
30. Gu Z, Zhang J, Liu L, et al. Rethinking reverse distillation for multi-modal anomaly detection. *Proc AAAI Conf Artif Intell* 2024; 38: 8445–8453.
31. Sun Z, Li X, Li Y, et al. Memoryless multimodal anomaly detection via student-teacher network and signed distance learning. In: *Pattern recognition and computer vision: 7th Chinese conference, PRCV 2024, Urumqi, China, 18–20 October 2024, proceedings, Part XII*, pp.447–461. Berlin, Heidelberg: Springer-Verlag. ISBN 978-981-97-8857-6. https://doi.org/10.1007/978-981-97-8858-3_31.
32. Zhou Z, Wang L, Fang N, et al. R3D-AD: reconstruction via diffusion for 3D anomaly detection. In: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T and Varol G (eds) *Computer vision – ECCV 2024: 18th European conference, Milan, Italy, 29 September–4 October 2024, proceedings, Part XXXVI*. Lecture Notes in Computer Science, volume 15094. Cham: Springer, pp.91–107. ISBN 978-3-031-72763-4. https://doi.org/10.1007/978-3-031-72764-1_6.
33. Chen R, Xie G, Liu J, et al. EasyNet: an easy network for 3D industrial anomaly detection. In: *Proceedings of the 31st ACM international conference on multimedia (MM '23)*. New York, NY, USA: Association for Computing Machinery, pp.7038–7046. ISBN 979-8-4007-0108-5. https://doi.org/10.1145/3581783.3611876.
34. Zavrtanik V, Kristan M and Skocaj D. Keep DRAEMing: discriminative 3D anomaly detection through anomaly simulation. *Pattern Recognit Lett* 2024; 181: 113–119.
35. Bi C, Li Y and Luo H. Dual-branch reconstruction network for industrial anomaly detection with RGB-D data. In: Pachori RB and Chen L (eds) *International conference on image, signal processing, and pattern recognition (ISPP 2024)*. Vol. 13180, p.1318033. International Society for Optics and Photonics, SPIE. https://doi.org/10.1117/12.3033181.
36. Zavrtanik V, Kristan M and Skocaj D. Cheating depth: enhancing 3D surface anomaly detection via depth simulation. In: *Proceedings of the IEEE/CVF winter conference on computer vision (WACV)*, pp.2153–2161. https://doi.org/10.1109/WACV57701.2024.00216.
37. Ye J, Zhao W, Yang X, et al. PO3AD: predicting point offsets toward better 3D point cloud anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.1353–1362. https://doi.org/10.1109/CVPR52734.2025.00134.
38. Chu YM, Liu C, Hsieh TI, et al. Shape-guided dual-memory learning for 3D anomaly detection. In: *Proceedings of the 40th international conference on machine learning (ICML)*, pp.6185–6194.
39. Horwitz E and Hoshen Y. 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 workshops (CVPRW)*, pp.2968–2977. https://doi.org/10.1109/CVPRW59228.2023.00298.
40. Cao Y, Xu X and Shen W. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognit* 2024; 156: 110761.
41. Wang Y, Peng J, Zhang J, et al. Multimodal industrial anomaly detection via hybrid fusion. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.8032–8041. https://doi.org/10.1109/CVPR52729.2023.00776.
42. Liu J, Mou S, Gaw N, et al. Uni-3DAD: Gan-inversion aided universal 3D anomaly detection on model-free products. *Expert Syst Appl* 2025; 272: 126665.
43. Tu Y, Zhang B, Liu L, et al. Self-supervised feature adaptation for 3D industrial anomaly detection. In: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T and Varol G (eds) *Computer vision – ECCV 2024*. Cham: Springer Nature Switzerland, pp.75–91. ISBN 978-3-031-72627-9. https://doi.org/10.1007/978-3-031-72627-9_5.
44. Roth K, Pemula L, Zepeda J, et al. Towards total recall in industrial anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.14318–14328. IEEE. https://doi.org/10.1109/CVPR52688.2022.01396.
45. Kruse M, Rudolph M, Woiwode D, et al. SplatPose & Detect: pose-agnostic 3D anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops*, pp.3950–3960. IEEE.
46. Jiang B, Xie Y, Li J, et al. IGSPAD: inverting 3D gaussian splatting for pose-agnostic anomaly detection. In: *Proceedings of the ACM international conference on multimedia (ACM MM)*.
47. Liu Y, Hu YS, Chen Y, et al. SplatPose+: real-time image-based pose-agnostic 3D anomaly detection. In: *Computer vision – ECCV 2024 workshops, Milan, Italy, 29 September–*

4 October 2024, proceedings, Part IV, pp.378–391. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-031-92804-8. https://doi.org/10.1007/978-3-031-92805-5_24.

48. Hartley R and Zisserman A. *Multiple view geometry in computer vision*. Cambridge, England: Cambridge University Press, 2004.
49. Snavely N, Seitz SM and Szeliski R. Photo tourism: Exploring photo collections in 3D. *ACM Trans Graph* 2006; 25: 835–846.
50. Sarlin PE, Cadena C, Siegwart R, et al. From coarse to fine: robust hierarchical localization at large scale. In: *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp.12708–12717. IEEE. https://doi.org/10.1109/CVPR.2019.01300.
51. Zheng P, Gao D, Fan DP, et al.Bilateral reference for high-resolution dichotomous image segmentation. *CAAI Artif Intell Res* 2024; 3: 9150038.
52. Suvorov R, Logacheva E, Mashikhin A, et al. Resolution-robust large mask inpainting with Fourier convolutions. In: *2022 IEEE/CVF winter conference on applications of computer vision (WACV)*, pp.3172–3182. IEEE. https://doi.org/10.1109/WACV51458.2022.00323.
53. Arandjelović R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition. *IEEE Trans Pattern Anal Mach Intell* 2018; 40: 1437–1451.
54. DeTone D, Malisiewicz T and Rabinovich A. SuperPoint: self-supervised interest point detection and description. In: *2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*, pp.337–349. IEEE. https://doi.org/10.1109/CVPRW.2018.00060.
55. Lindenberger P, Sarlin PE and Pollefeys M. LightGlue: local feature matching at light speed. In: *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp.17581–17592. IEEE. https://doi.org/10.1109/ICCV51070.2023.01616.
56. Tan M and Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri K and Salakhutdinov R (eds.) *Proceedings of the 36th international conference on machine learning (ICML)*. Proceedings of Machine Learning Research, Volume 97, pp.6105–6114. PMLR.
57. Blender Online Community. *Blender: a 3D modelling and rendering package*, 2018.