A Survey on Industrial Anomaly Synthesis

Anonymous authors
Paper under double-blind review

Abstract

This paper presents a comprehensive review of industrial anomaly synthesis (IAS). Existing surveys on industrial anomalies mainly focus on anomaly detection, while anomaly synthesis is typically treated as an auxiliary component rather than as an independent topic. However, owing to its increasing importance in data augmentation, downstream model training, and controllable industrial inspection, IAS has become a research direction of growing interest. To address the lack of a dedicated review, we survey a broad range of representative methods and organize them into four paradigms: hand-crafted synthesis, distribution hypothesis-based synthesis, generative model (GM)-based synthesis, and vision-language model (VLM)-based synthesis. We further establish a dedicated taxonomy for IAS, which supports more systematic comparison across methods and offers a clearer view of the field's development. Beyond methodological categorization, we summarize the datasets, benchmarks, and evaluation metrics commonly adopted in IAS, and review recent advances in multimodal anomaly synthesis that remain underexplored in prior surveys. Overall, this survey provides a structured understanding of existing IAS methods, evaluation settings, current limitations, and promising future directions, and is intended to serve as a reference for subsequent research in this area. More resources are available at https://anonymous.4open.science/status/IAS.

1 Introduction

Image anomaly detection plays an important role in manufacturing because it helps identify abnormal products and thereby supports product quality, production safety, and process reliability. In practical industrial settings, however, building effective image anomaly detection systems usually relies on a sufficient number of high-quality annotated abnormal samples for training, and obtaining such samples is often costly and difficult. The main reasons are summarized as follows: (1) The proportion of abnormal products is usually very low in large-scale manufacturing, resulting in a natural scarcity of real abnormal samples. (2) Many industrial anomaly patterns, such as microscopic cracks, fine scratches, concealed contaminants, or internal structural anomalies, require specialized inspection equipment, *e.g.*, high-magnification microscopes, X-ray systems, or infrared devices, which significantly increases the cost of data acquisition. (3) High-quality annotation further requires domain expertise and careful analysis. In many cases, skilled professionals are needed to identify abnormal regions accurately, and some samples may even require multimodal annotation, which further increases the time and labor cost.

To alleviate the shortage of real abnormal samples, a growing number of industrial anomaly synthesis (IAS) methods have been developed for data augmentation and downstream model training. As shown in Fig. 1, IAS has attracted rapidly increasing attention in recent years, reflecting its growing role in industrial inspection. This trend is closely related to practical demand. In many real applications, it is often not sufficient to only determine whether a sample is abnormal. Instead, practitioners increasingly expect IAS to support diverse anomaly synthesis, segmentation-oriented supervision, and more controllable training data construction under specific industrial contexts. IAS is also expected to facilitate benchmark construction, stress testing, and failure-case analysis for downstream inspection systems. In this sense, IAS is gradually evolving from a simple augmentation strategy into a more task-oriented component for industrial model development and evaluation. Nevertheless, despite this progress, current IAS methods still cannot fully satisfy practical industrial requirements. Their main limitations can be summarized into three aspects:

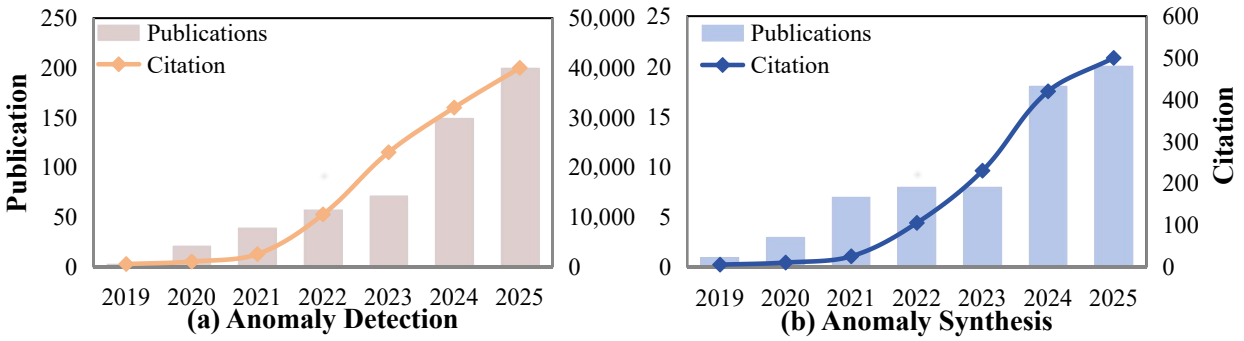

Figure 1: Trend of papers related to anomaly detection and anomaly synthesis from 2019 to 2025. Anomaly detection has shown steady growth in publications, while anomaly synthesis has attracted increasing attention in recent years, with a clear surge in 2025. Although many surveys have reviewed anomaly detection, anomaly synthesis remains an emerging topic with limited dedicated review efforts.

(1) **Limited Coverage of Abnormal Distribution.** In practical industrial scenarios, the number of available abnormal samples for a specific anomaly type is often very small. As a result, existing methods usually capture only a limited part of the underlying anomaly distribution, which restricts diversity and weakens generalization. This issue is particularly evident for long-tail or highly irregular anomalies, such as subtle cracks, local corrosion, stains, punctures, or composite anomalies that vary greatly in shape, scale, and texture.

(2) **Difficulty in Synthesizing Realistic Abnormal Samples.** Real industrial abnormal patterns are often structurally complex and visually diverse, while also being tightly coupled with local material properties, imaging conditions, and background context. For example, a fine scratch on a polished metal surface, a contamination spot on transparent packaging, or internal anomalies revealed by X-ray imaging may each exhibit very different visual characteristics. This makes it difficult for synthesis methods to preserve both realism and consistency. As a result, synthesized results may still contain unrealistic textures, missing details, or unnatural transition boundaries.

(3) **Limited Use of Multimodal Information for Controllable Synthesis.** In industrial scenarios, useful cues may come from multiple modalities, such as text descriptions, masks, spatial priors, reference images, or other auxiliary signals. These cues are important when users need to specify where an anomaly should appear, what type of anomaly should be synthesized, or how strong the abnormal pattern should be. However, how to effectively incorporate such multimodal information into IAS remains insufficiently explored, especially for controllable and realistic anomaly synthesis.

Although IAS has developed rapidly, a dedicated and systematic review remains lacking. Existing anomaly-related surveys are not completely unrelated to anomaly synthesis; many of them mention data augmentation, generative modeling, or synthetic anomalies within broader industrial anomaly detection pipelines. However, these discussions are usually embedded in detection-oriented narratives, where anomaly synthesis is treated as an auxiliary tool rather than as an independent methodological problem. As summarized in Table 1, prior surveys have provided valuable discussions on visual inspection, data augmentation, generative modeling, and industrial anomaly detection, but they do not yet offer a unified synthesis-centered understanding of IAS itself. In particular, current reviews usually cover only part of the synthesis literature, rarely organize methods according to their dominant anomaly-forming mechanisms, and provide limited discussion of the recent shift toward multimodal and VLM-based controllable synthesis. Consequently, it remains difficult to systematically compare how different IAS methods construct anomalies, what inputs and supervision signals they require, and how their outputs support downstream tasks. It also remains difficult to clearly trace how the field has evolved from early rule-based perturbation and latent-space deviation to recent image-space generation, local editing, and multimodal controllable synthesis.

Motivated by this gap, this paper presents a dedicated survey on industrial anomaly synthesis. Rather than discussing anomaly synthesis only as a supporting tool for anomaly detection, we study IAS as an

Table 1: Comparison of previous surveys and our survey.

Ref.	Year	Perspective	IAS Cat.	Multimodal	Description
Chen et al. (2021)	2021	Detection	0	✗	Visual inspection review with emphasis on augmentation and practical pipelines.
Xia et al. (2022)	2022	Detection	4	✗	GAN-centric review covering GAN-based techniques related to AD and synthesis.
Liu et al. (2024)	2024	Detection	2	✗	Broad anomaly detection survey; anomaly synthesis is not a primary focus.
Cao et al. (2024)	2024	Detection	4	✗	Discusses anomaly-related topics and mentions multimodal cues; IAS is not systematically reviewed.
Mao et al. (2025)	2025	Detection	1	✗	Industrial anomaly detection survey covering paradigms, benchmarks, and metrics; includes synthesis-related discussion.
Li et al. (2025)	2025	Detection	1	✗	Industrial visual anomaly detection survey covering learning strategies, generative modeling, and multimodal/VLM-based detection methods.
Ours	**2026**	**Synthesis**	**10**	✓	**Dedicated IAS taxonomy; reviews ∼60 representative IAS methods across paradigms; summarizes benchmarks and practical trade-offs.**

independent methodological direction with its own synthesis mechanisms, benchmark choices, evaluation settings, and practical trade-offs. To this end, we organize existing IAS methods into four paradigms, namely hand-crafted synthesis, distribution hypothesis-based synthesis, generative model (GM)-based synthesis, and vision-language model (VLM)-based synthesis. We first summarize commonly used datasets and evaluation protocols, and then review representative IAS methods from both methodological and practical perspectives based on the proposed taxonomy. We also discuss the current limitations, open challenges, and promising future directions of IAS.

Overall, the main contributions of this paper are summarized as follows:

- We present a dedicated taxonomy for industrial anomaly synthesis (IAS), which organizes existing methods into four paradigms: hand-crafted synthesis, distribution hypothesis-based synthesis, generative model (GM)-based synthesis, and vision-language model (VLM)-based synthesis. This taxonomy provides a structured and fine-grained framework for understanding methodological differences, technical evolution, and practical design choices in IAS.

- We summarize commonly used datasets, benchmarks, and evaluation protocols, thereby clarifying how IAS methods are commonly developed and assessed in the literature. Based on the proposed taxonomy, we then provide a unified and systematic review of a broad range of representative IAS methods.

- We discuss recent advances, current limitations, open challenges, and promising future directions of IAS research, with particular attention to controllability, semantic alignment, multimodal conditions, and industrial applicability.

The remainder of this paper is structured as follows: Section 2 first introduces commonly used datasets, benchmarks, and evaluation protocols in IAS, and then presents the proposed taxonomy with its key paradigms and methodological distinctions. Sections 3 to 6 provide an in-depth analysis of the four primary categories of IAS, including Hand-crafted synthesis, Distribution hypothesis-based synthesis, GM-based synthesis, and recently emerging VLM-based synthesis. Finally, Sections 7 and 8 summarize key insights from the survey, discuss the limitations of current methods, and outline promising future research directions for IAS.

2 Taxonomy and Benchmarking Overview

2.1 Taxonomy and Definitions of IAS

Fig. 2 presents the overall taxonomy of industrial anomaly synthesis (IAS). In this survey, IAS methods are organized into four main paradigms, namely hand-crafted synthesis, distribution hypothesis-based synthesis, GM-based synthesis, and VLM-based synthesis. The classification follows the dominant synthesis mechanism, that is, the stage that primarily forms the anomaly and defines the corresponding controllability for downstream use. These paradigms differ in their synthesis mechanisms, degree of controllability, and applicable scenarios. To further illustrate their internal logic, Figs. 3–6 provide paradigm-specific visual summaries of how anomaly synthesis is conducted in representative approaches.

Hand-crafted synthesis relies on manually designed rules to construct samples with anomalies. It is usually training-free and is suitable for controlled settings where high realism and large anomaly diversity are not the primary requirements. Within this paradigm, self-contained synthesis directly manipulates the original image through cropping, copying, rearrangement, or related operations to form abnormal regions. External-dependent synthesis introduces auxiliary sources such as texture libraries, so that the synthesized content is not restricted to the image itself. Inpainting-based synthesis first masks local regions and then disrupts structural completeness by inserting noise, black patches, or missing content. Because the abnormal content is not generated from an explicitly learned anomaly distribution, these methods are usually simple and effective for data expansion, but their realism and distributional coverage are often limited.

Distribution hypothesis-based synthesis generates anomalies by modeling normal data distributions and introducing controlled deviations, typically in latent space or feature space. Prior-dependent methods adopt predefined geometric assumptions, such as manifold or hypersphere structures, to characterize normality, and then synthesize abnormal features near or beyond the normal boundary. By contrast, data-driven methods do not rely on explicit geometric priors. Instead, they exploit intrinsic statistical properties of the data and synthesize anomalies through perturbation or adaptive feature manipulation, which often provides greater flexibility and diversity. As a result, this paradigm is often attractive for feature-level training augmentation and anomaly-aware representation learning, although the synthesized results may be less intuitive at the pixel level than image-space generation methods.

GM-based synthesis uses deep generative models, such as GANs and diffusion models, to produce more realistic abnormal samples. This paradigm can be further divided into full-image synthesis, full-image translation, and local anomaly synthesis. Full-image synthesis learns an abnormal distribution and maps random noise to abnormal samples. Full-image translation transforms normal images into abnormal ones while preserving the overall scene structure. Local anomaly synthesis modifies selected regions of normal images with generated abnormal content and emphasizes local consistency between the edited area and its surrounding background. Compared with earlier paradigms, GM-based methods generally offer stronger visual realism and richer anomaly appearance, but they also tend to require heavier training, more careful optimization, and higher computational cost.

VLM-based synthesis leverages large-scale pre-trained vision-language models together with multimodal conditions to produce high-quality abnormal samples. Single-stage methods directly synthesize context-aware anomalies, usually through prompt engineering or lightweight fine-tuning. Multi-stage methods instead adopt a staged pipeline that combines anomaly generation with additional processes such as mask generation, refinement, or sequential optimization, thereby improving realism, diversity, and downstream compatibility. Their main advantage lies in stronger semantic controllability and better use of multimodal cues, especially when anomaly type, location, or contextual compatibility needs to be specified more explicitly.

Although these four paradigms are presented as separate branches, they are not completely isolated in practice. Hand-crafted and distribution hypothesis-based methods usually emphasize efficiency and training utility, whereas GM-based and VLM-based methods place more emphasis on realism, spatial control, and annotation compatibility. The taxonomy therefore also provides a basis for comparing methods from both methodological and practical perspectives.

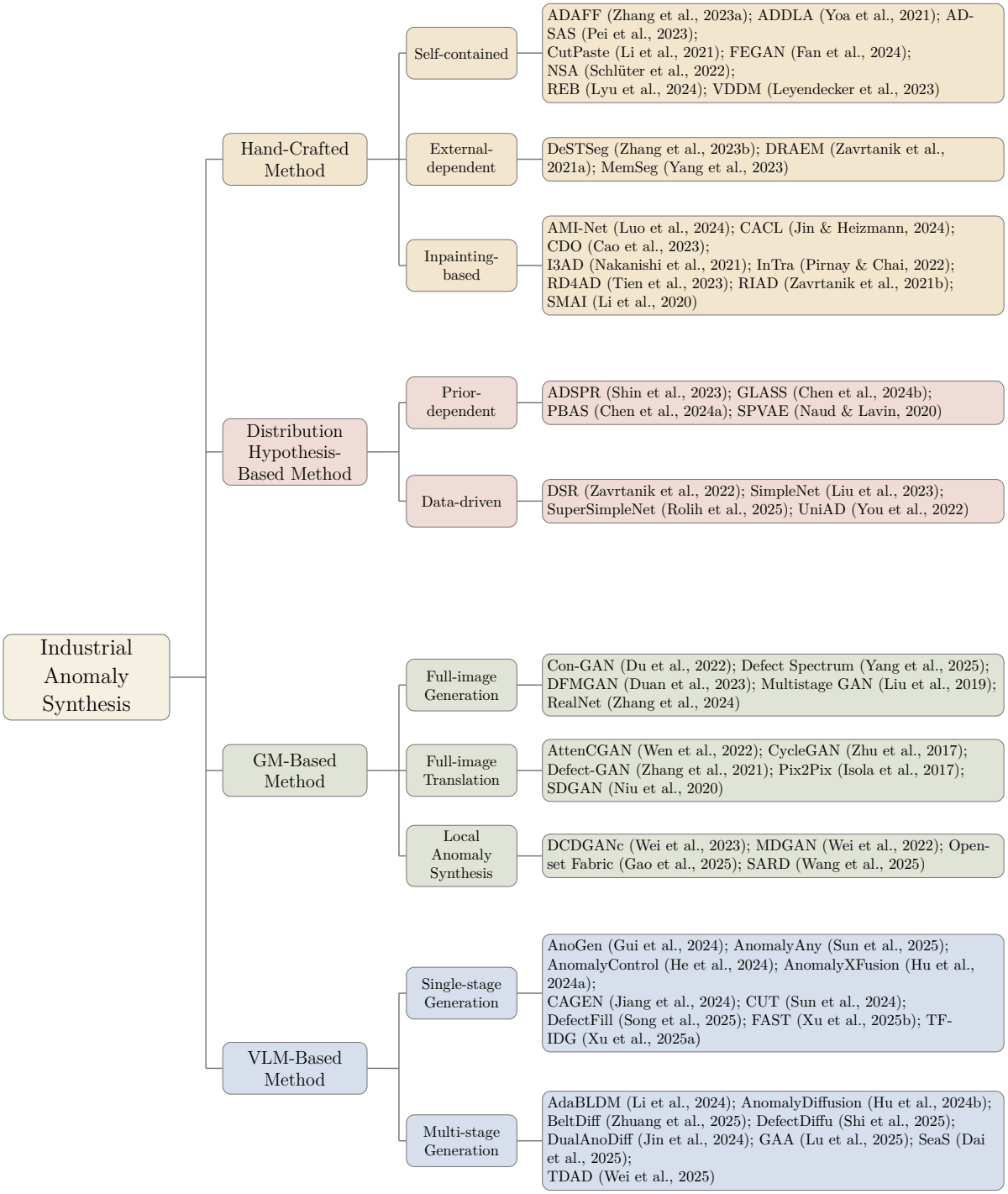

Figure 2: A taxonomy of industrial image anomaly synthesis (IAS).

2.2 Datasets, Benchmarks, and Evaluation Protocols

IAS results are not directly comparable unless benchmark choice, supervision setting, and evaluation context are considered together.

Table 2: Representative datasets commonly used in IAS-related studies. "Data type" indicates whether the benchmark is primarily based on real industrial imagery (Real), fully synthetic data (Syn.), or a mixture of real and synthetic components (Mixed).

Dataset	Year	Venue	Data type (Real/Syn./Mixed)	Samples	Classes	IAS relevance				Site
						H	DH	GM	VLM	
MVTec-AD Bergmann et al. (2019)	2019	CVPR	Real	5,354	15	✓	✓	✓	✓	link
VisA Zou et al. (2022)	2022	ECCV	Real	10,821	12	✓	✓	✓	✓	link
BTAD Mishra et al. (2021)	2021	ISIE	Real	2,830	3	✓	✓	✓	✗	link
MPDD Jezek et al. (2022)	2022	ICUMT	Real	>1,000	6	✓	✓	✓	✗	link
KSDD Tabernik et al. (2020)	2020	JIM	Real	399	1	✓	✓	✓	✗	link
KSDD2 Božič et al. (2021)	2021	Comput. Ind.	Real	3,335	1	✓	✓	✓	✗	link
AITEX Silvestre-Blanes et al. (2019)	2019	Autex Res. J.	Real	245	7	✓	✓	✓	✗	link
MVTec LOCO-AD Bergmann et al. (2022a)	2022	IJCV	Real	3,644	5	✗	✓	✓	✓	link
PCB-Bank Yao et al. (2024)	2024	ECCV	Real	6,749	7	✗	✓	✓	✓	link
Real-IAD Wang et al. (2024)	2024	CVPR	Real	150,000+	30	✗	✗	✓	✗	link
MVTec 3D-AD Bergmann et al. (2022b)	2022	VISAPP	Real	4,147	10	✗	✗	✓	✓	link
MVTec AD 2 Heckler-Kram et al. (2025)	2025	arXiv	Real	8,000+	8	✗	✗	✓	✓	link
DAGM 2007 Wieler et al. (2007)	2007	DAGM	Syn.	16,100	10	✓	✗	✓	✗	link
DTD-Synthetic Aota et al. (2023)	2023	WACV	Syn.	>2,400	12	✓	✓	✓	✗	link
Eyecandies Bonfiglioli et al. (2022)	2022	ACCV	Syn.	90,000	10	✗	✗	✓	✓	link
PAD Zhou et al. (2023)	2023	NeurIPS	Mixed	11,000+	20	✗	✗	✓	✗	link
ISP-AD Krassnig & Gruber (2025)	2025	arXiv	Mixed	559,049	3	✗	✗	✓	✗	link

H = Hand-crafted, DH = Distribution hypothesis-based, GM = generative model-based, VLM = vision-language model-based. IAS relevance indicates representative, non-exhaustive usage in reviewed studies.

Table 2 summarizes representative benchmarks used in IAS-related studies, covering real, synthetic, and mixed data sources. Rather than differing only in data type, these datasets also vary in scale, inspection setting, and annotation characteristics. Among real-world benchmarks, MVTec-AD Bergmann et al. (2019) and VisA Zou et al. (2022) are widely adopted as standard references for 2D industrial anomaly detection. Other datasets, such as BTAD Mishra et al. (2021), MPDD Jezek et al. (2022), KSDD Tabernik et al. (2020), and KSDD2 Božič et al. (2021), are more closely associated with specific industrial inspection scenarios and typically involve more constrained object categories or acquisition conditions. MVTec LOCO-AD Bergmann et al. (2022a) extends this setting by including logical anomalies in addition to structural anomalies, while Real-IAD Wang et al. (2024) introduces large-scale, high-resolution, and multi-view data collected from real production lines. The datasets also differ in sensing modality and construction strategy. MVTec 3D-AD Bergmann et al. (2022b) focuses on industrial 3D sensing, whereas Eyecandies Bonfiglioli et al. (2022) provides a synthetic multimodal setting with RGB, depth, and normal maps. In contrast, datasets such as DAGM Wieler et al. (2007) and DTD-Synthetic Aota et al. (2023) are fully synthetic and mainly used for controlled or texture-oriented evaluation. Mixed datasets, including PAD Zhou et al. (2023) and ISP-AD Krassnig & Gruber (2025), combine real and synthetic components to balance controllability and realism. Overall, these variations imply that IAS methods may be evaluated under different assumptions depending on the selected benchmark, including differences in visual appearance, anomaly type, and supervision setting.

Metric selection is therefore not a neutral reporting detail. Some studies mainly test whether synthesized anomalies improve downstream detectors, whereas others further examine visual realism, structural preservation, editing faithfulness, or image–mask alignment. Table 3 organizes these criteria into synthesis/editing, downstream detection, and localization/segmentation metrics, which correspond to different technical claims in IAS evaluation.

Table 3: Key metrics commonly used in IAS evaluation.

Metric	Level	Formula	Remarks/usage
Synthesis / Editing Metrics			
FID	↓	$\|\mu_r - \mu_g\|_2^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}\right)$	Distributional realism between real and generated features (often computed on full images or cropped regions / patches).
LPIPS	↓	$d(x, y)$	Perceptual distance; commonly used for visual fidelity and sometimes diversity.
PSNR	↑	$10\log_{10}\left(\mathrm{MAX}^2/\mathrm{MSE}\right)$	Pixel-level reconstruction fidelity; often used to assess background preservation in local editing.
SSIM	↑	$\frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$	Structure preservation under synthesis or editing.
Alignment IoU	↑	$\frac{\lvert\hat{M} \cap M\rvert}{\lvert\hat{M} \cup M\rvert}$	Mask-conditioned control fidelity (generated / edited region vs. target mask).
Downstream Detection Metrics			
Precision (P)	↑	$\frac{TP}{TP + FP}$	Controls false alarms; proportion of predicted anomalies that are correct.
Recall (R)	↑	$\frac{TP}{TP + FN}$	Coverage of true anomalies; avoids missed anomalies.
FPR	↓	$\frac{FP}{FP + TN}$	False alarm rate on normal samples; often used to trace ROC / PRO-style curves.
Image-level AUROC	↑	$\int_0^1 \mathrm{TPR}(u)\, du, \quad u = \mathrm{FPR}$	Threshold-free image-level anomaly detection performance.
Image-level AP	↑	$\sum_n (R_n - R_{n-1}) P_n$	Preferred under class imbalance; summarizes the precision–recall curve without linear interpolation.
F1 score	↑	$\frac{2PR}{P + R}$	Balances precision and recall; overall image-level detection effectiveness.
Localization / Segmentation Metrics			
Pixel-level AUROC	↑	$\int_0^1 \mathrm{TPR}(u)\, du, \quad u = \mathrm{FPR}$	Pixel-wise localization quality (anomaly map vs. GT mask).
PRO	↑	$\frac{1}{\lvert\mathcal{C}\rvert} \sum_{C \in \mathcal{C}} \frac{\lvert\hat{M}_\tau \cap C\rvert}{\lvert C\rvert}$	Per-region overlap at threshold τ; averages region-wise recall over connected GT components.
AU-PRO	↑	$\int_0^\alpha \mathrm{PRO}(u)\, du, \quad u = \mathrm{FPR}$	Area under the PRO curve; commonly reported with an FPR limit α (often 0.3).
sPRO	↑	$\frac{1}{\lvert\mathcal{C}\rvert} \sum_{C \in \mathcal{C}} \min\left(\frac{\lvert\hat{M}_\tau \cap C\rvert}{s(C)}, 1\right)$	Saturated PRO for settings with logical / weakly pixel-defined anomalies; $s(C)$ is a dataset-defined saturation threshold for region C.
AU-sPRO	↑	$\int_0^\beta \mathrm{sPRO}(u)\, du, \quad u = \mathrm{FPR}$	Area under the sPRO curve; commonly used in MVTec LOCO-style evaluation, often with $\beta = 0.05$.
IoU	↑	$\frac{\lvert\hat{M} \cap M\rvert}{\lvert\hat{M} \cup M\rvert}$	Thresholded mask overlap for segmentation / localization.
mIoU	↑	$\frac{1}{N} \sum_{i=1}^{N} \frac{\lvert\hat{M}_i \cap M_i\rvert}{\lvert\hat{M}_i \cup M_i\rvert}$	Mean IoU over samples.
Dice (mask-F1)	↑	$\frac{2\lvert\hat{M} \cap M\rvert}{\lvert\hat{M}\rvert + \lvert M\rvert}$	Segmentation overlap; often more sensitive to small anomalous regions than IoU.

TP, FP, TN, FN denote true / false positives / negatives; $\hat{M}$ and M denote predicted and ground-truth masks; $\hat{M}_\tau$ denotes the thresholded predicted mask at threshold τ; $\mathcal{C}$ is the set of connected components in the ground-truth anomaly mask; $s(C)$ denotes the saturation threshold associated with region C; α and β are the integration limits used for AU-PRO and AU-sPRO, respectively.

Metric design should also be interpreted under the same benchmark logic. As the dataset family, anomaly form, and supervision granularity change, the role of evaluation metrics changes accordingly. This distinction is important because realism, detector improvement, and spatial accuracy do not correspond to the same notion of synthesis quality. Accordingly, IAS evaluation can be understood from three complementary perspectives: direct synthesis quality, downstream detection utility, and spatial localization accuracy.

The first perspective concerns direct synthesis quality. Metrics such as FID and LPIPS are most informative when the generated image is directly used as an evaluation target rather than merely as a training artifact. In this setting, the question is not only whether the synthesized anomaly helps detection, but also whether

it is visually plausible and statistically close to real abnormal data. FID is commonly used as a coarse feature-space realism indicator, whereas LPIPS is more sensitive to perceptual similarity. When the task involves local editing rather than full-image generation, PSNR and SSIM become more meaningful because they reflect background preservation and structural continuity outside the manipulated region. Alignment IoU is particularly relevant in mask-guided settings, where a method is expected to place the anomaly in the intended region with sufficient spatial controllability.

The second perspective concerns downstream detection utility. IAS is rarely judged by visual quality alone, because many studies use synthesized anomalies primarily to improve anomaly detection through more effective training data. Image-level AUROC remains one of the most widely used summary metrics because it is threshold-free and relatively easy to compare across methods. However, AUROC alone may be insufficient in industrial settings with low anomaly prevalence and substantial class imbalance. Under such conditions, image-level AP is often more informative because it captures the precision–recall trade-off more directly, while F1 score is useful for reporting balanced performance at a selected operating point. Precision and recall further help clarify whether a reported gain mainly comes from better anomaly coverage, fewer false alarms, or a compromise between the two. Together, these metrics address a central practical question in IAS: whether synthesized anomalies genuinely improve downstream discrimination rather than merely producing visually appealing abnormal samples.

The third perspective concerns spatial localization accuracy. Once a method claims stronger spatial correspondence, region-aware supervision, or improved local controllability, localization and segmentation metrics become necessary. Pixel-level AUROC is a common starting point because it evaluates anomaly maps in a threshold-free manner, but pixel-wise ranking alone does not fully characterize localization quality in industrial settings. PRO and AU-PRO are therefore particularly important, since they assess overlap at the connected-region level and are often better aligned with practical anomaly localization than raw pixel accuracy alone. IoU, mIoU, and Dice are more directly tied to explicit mask agreement and are especially useful for local synthesis, mask-guided editing, or paired supervision. Dice is often more sensitive to small abnormal regions, whereas IoU imposes a stricter overlap criterion. In settings involving logical anomalies or less sharply defined abnormal regions, sPRO and AU-sPRO become more appropriate because they better match the evaluation protocols used in such settings.

These three perspectives are not equally emphasized across IAS paradigms. For hand-crafted and distribution hypothesis-based methods, synthesized anomalies often function mainly as intermediate training resources, so the most direct evidence of usefulness usually comes from downstream detection gains. By contrast, GM-based and VLM-based methods often make stronger claims about realism, controllability, image–mask consistency, or prompt-guided editing quality, which cannot be supported by AUROC or AP alone and therefore require direct synthesis metrics and alignment-oriented measures as well. When such methods are further used to provide dense supervision or spatially controllable augmentation, localization metrics are needed to verify that improved realism does not come at the expense of accurate region correspondence.

Metric choice is therefore tightly coupled with benchmark choice. On standard real 2D datasets such as MVTec-AD and VisA, image-level AUROC, pixel-level AUROC, AU-PRO, IoU, and Dice often provide a sufficiently solid picture of whether synthesized anomalies improve detector training and localization quality. On MVTec LOCO-AD, protocol-aware measures such as sPRO or AU-sPRO become more suitable because the evaluation target includes logical inconsistency in addition to ordinary abnormal regions. In local editing or controllable generation settings, metrics such as SSIM, PSNR, LPIPS, and Alignment IoU carry greater weight because they directly reflect whether the generated anomaly is visually and spatially well grounded. Metric selection in IAS should therefore be understood as part of the technical claim itself rather than as a cosmetic reporting choice.

Taken together, Table 3 provides a compact view of how IAS methods are evaluated in practice. Direct synthesis metrics indicate whether the generated anomaly is visually plausible and aligned with the intended control signal; downstream detection metrics indicate whether the synthetic data improves discriminative learning; and localization metrics indicate whether this benefit extends to spatially precise anomaly identification. Read together with Table 2, this metric view establishes the evaluation framework for the cross-paradigm review that follows.

3 Hand-crafted Synthesis

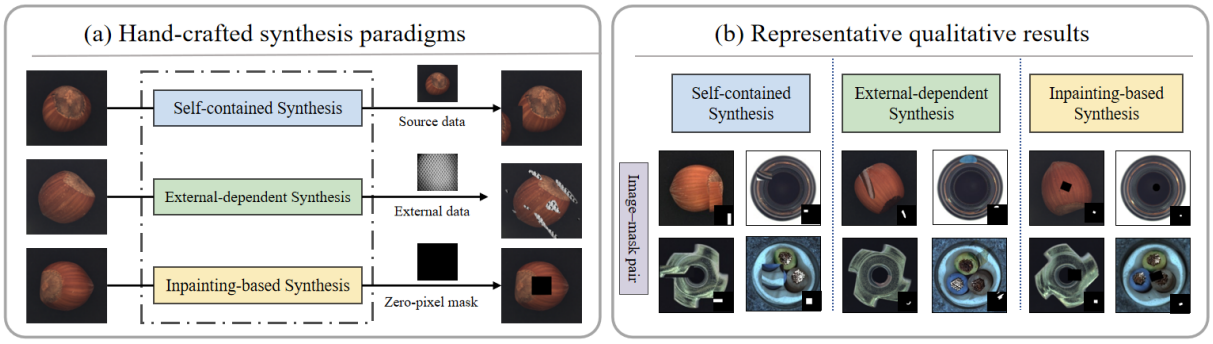

Figure 3: Hand-crafted synthesis. (a) Overview of three representative hand-crafted pipelines, including self-contained synthesis, external-dependent synthesis, and inpainting-based synthesis. These methods synthesize anomaly content through rule-based image manipulation without training a dedicated generative model. (b) Representative qualitative results adapted from Xu et al. (2025b;c), including examples of methods such as CutPaste Li et al. (2021) and DRAEM Zavrtanik et al. (2021a).

As illustrated in Fig. 3, hand-crafted synthesis constructs samples with anomalies through predefined operations on normal images. Compared with later paradigms that rely on learned generative priors, this family is training-free, simple to implement, and easy to deploy in low-resource settings. Its central idea is to create anomalies by manually altering image content or visual completeness, and the main differences among subcategories lie in whether the anomaly is derived from the image itself, introduced from external sources, or formed by deliberately disrupting local structure.

Self-contained synthesis represents the most direct form of hand-crafted IAS, because the synthesized abnormal content is derived entirely from the original image itself. Typical operations include cropping, copying, rearranging, and local perturbation. Early representative methods such as CutPaste Li et al. (2021) and NSA Schlüter et al. (2022) randomly crop patches and paste them back into normal images, while REB Lyu et al. (2024) introduces Bézier-curve-based region design to obtain more flexible and irregular anomaly shapes. Other methods further extend this line by using traditional augmentation, local copying, or region-level transformation to perturb the original image more adaptively, as in ADSAS Pei et al. (2023), ADAFF Zhang et al. (2023a), VDDM Leyendecker et al. (2023), and FEGAN Fan et al. (2024). ADDLA Yoa et al. (2021) further combines multiple local transformations dynamically, making the synthesis process less rigid than simple crop-and-paste designs. Because these methods operate within native image content, they usually preserve structural and textural consistency better than other hand-crafted variants. This makes them attractive for rapid augmentation and for scenarios where contextual coherence is more important than anomaly novelty. However, the synthesized anomalies are still constrained by the diversity of the source image itself. As a result, they often struggle to approximate more complex real anomalies whose appearance cannot be easily simulated through internal rearrangement alone.

External-dependent synthesis builds on the previous idea by introducing anomaly content that does not originate from the source image itself. Instead of relying only on internal structures, it blends external textures or auxiliary visual patterns into normal images, thereby increasing appearance diversity. Representative methods such as DRAEM Zavrtanik et al. (2021a) and DeSTSeg Zhang et al. (2023b) use Perlin-noise-based masks to combine clean backgrounds with external textures, while MemSeg Yang et al. (2023) further constrains the anomaly region to the foreground to reduce semantic mismatch between synthesized anomalies and object regions. Relative to self-contained synthesis, this subcategory broadens the anomaly space by allowing content that is unavailable in the original image. Its main advantage is that it can generate richer appearance variation than purely internal manipulation, which is particularly useful when the target anomaly patterns differ substantially from the texture statistics of normal samples. At the same time, once anomaly content is imported from external sources, local compatibility becomes more difficult to maintain. Blending

artifacts, boundary inconsistency, and semantic mismatch may therefore appear, especially when the inserted texture is visually plausible in isolation but poorly aligned with the surrounding industrial context.

Table 4: Practical profile of hand-crafted IAS.

Family	**Resource burden**			**Practical ceiling**	
	Train	**Infer**	**Memory**	**Control**	**Realism**
Self-contained	–	○	○	◐	○
External-dependent	–	○	◐	◐	◐
Inpainting-based	–	○	○	◐	○

○/◐/●= low/medium/high; –= not applicable. Train/Infer/Memory denote relative training cost, inference cost, and memory footprint, respectively.

Inpainting-based synthesis differs from the previous two subcategories by creating anomalies through masking selected regions and intentionally breaking local visual completeness. In this sense, it can be viewed as a complementary hand-crafted strategy: whereas self-contained and external-dependent methods mainly create "added" abnormal content, inpainting-based methods more often simulate missing, corrupted, or occluded patterns. This makes them particularly compatible with reconstruction-oriented anomaly detection. Typical methods such as I3AD Nakanishi et al. (2021), RIAD Zavrtanik et al. (2021b), and SMAI Li et al. (2020) generate pseudo-anomalies by masking random regions and reconstructing them from surrounding information. Later variants improve this basic recipe by making masking more adaptive or structured. For example, InTra Pirnay & Chai (2022) introduces transformer-based inpainting for anomaly-oriented reconstruction, and AMI-Net Luo et al. (2024) develops an adaptive mask generator that preserves surrounding normal background while selectively concealing target regions. Other methods, including CDO Cao et al. (2023), RD4AD Tien et al. (2023), and CACL Jin & Heizmann (2024), further explore masking-based perturbations through random noise injection, cutout-style black masks, or related corruption strategies. This subcategory is simple, efficient, and especially suitable for reconstruction-style settings, because it directly constructs incomplete or corrupted local patterns that can challenge restoration-based models. It also avoids the need for external texture sources and usually remains easy to control spatially. However, its synthesized anomalies are often heuristic and structurally simple. When masking or corruption rules are overly crude, the generated anomalies may resemble artificial missing-content artifacts rather than realistic industrial anomalies with complex morphology.

Table 4 summarizes the practical characteristics of hand-crafted IAS from an engineering perspective. Since these methods are rule-based, their training cost is marked as –across all three subcategories, and inference overhead is generally low because synthesis mainly relies on patch manipulation, mask sampling, blending, or simple local corruption rather than iterative generation. The main runtime difference lies in memory usage: self-contained and inpainting-based methods typically require only lightweight buffers, whereas external-dependent synthesis may additionally maintain texture banks or reference sources. Overall, the three subcategories reflect different trade-offs within the same low-cost paradigm. Self-contained methods better preserve contextual consistency but are limited in anomaly diversity; external-dependent methods expand appearance variation but face greater local mismatch risk; and inpainting-based methods are well aligned with reconstruction-style pipelines, although their generated patterns often remain structurally simple. Therefore, hand-crafted synthesis is most suitable for rapid deployment, low-resource settings, and lightweight augmentation, while its limitations become more apparent when downstream tasks require high-fidelity textures, complex anomaly appearance, or strict image–mask consistency.

4 Distribution hypothesis-based Synthesis

As illustrated in Fig. 4, distribution hypothesis-based synthesis generates anomalies through controlled deviations in the latent or feature space of normal samples. Compared with hand-crafted synthesis, this paradigm no longer relies on direct pixel-space manipulation, but instead perturbs learned representations to simulate anomalies. It is usually more lightweight than image-space generative modeling and often effective for downstream detection or classification, yet its synthesized results are frequently less explicit in spatial detail. The

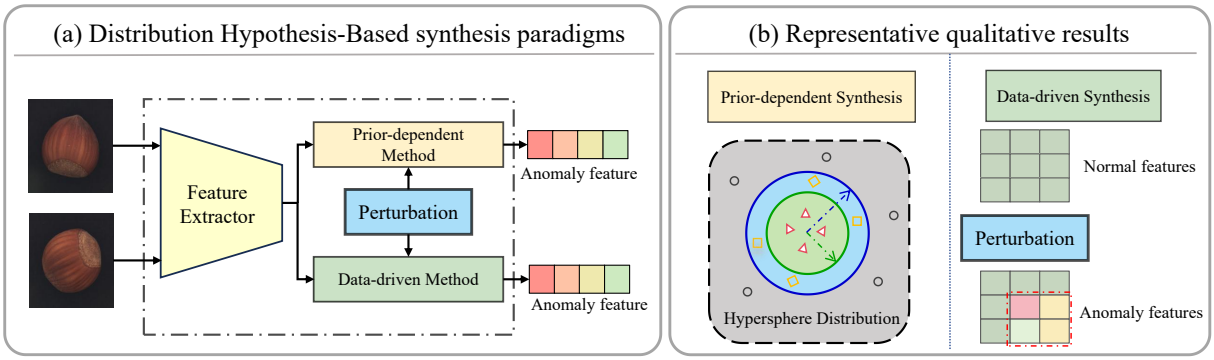

Figure 4: Distribution hypothesis-based synthesis. (a) Overview of the distribution hypothesis-based paradigm. It synthesizes abnormal samples by perturbing the learned feature or latent distribution of normal data, including prior-dependent strategies with explicit geometric assumptions and data-driven strategies with adaptive perturbations. (b) Representative qualitative illustrations adapted from Chen et al. (2024b).

main distinction within this paradigm lies in whether anomaly synthesis is guided by predefined geometric assumptions about the normal feature distribution or by data-driven perturbation strategies learned more directly from the data itself.

Prior-dependent synthesis represents the more structured branch of this paradigm, because it synthesizes anomalies under explicit assumptions about the geometry of the normal feature space. Typical formulations assume that normal samples lie within a compact manifold, hypersphere, or other constrained region, while abnormal features appear near or beyond its boundary. Representative methods such as GLASS Chen et al. (2024b) constrain normal features within a compact space and synthesize anomalies through gradient ascent together with truncated projection, while PBAS Chen et al. (2024a) uses progressive radial perturbation to refine the decision boundary. ADSPR Shin et al. (2023) further introduces score-based perturbation under manifold-style assumptions and a Gaussian annulus, and SPVAE Naud & Lavin (2020) explores anomaly synthesis in non-Euclidean latent spaces through geodesic perturbation. Because these methods explicitly model where normality resides in representation space, they can generate boundary-aware abnormal features in a relatively controlled manner. This makes them attractive for downstream detection settings in which the goal is to improve decision boundaries rather than to synthesize visually rich abnormal textures. However, their controllability is mainly expressed at the feature level rather than in image space. As a result, the synthesized anomalies usually lack explicit spatial structure and are less suitable for tasks that demand pixel-level realism or fine-grained segmentation supervision. In addition, once the assumed geometry of the normal distribution becomes too restrictive, these methods may struggle to capture more complex or irregular anomaly patterns.

Data-driven synthesis relaxes these explicit geometric assumptions and instead perturbs latent representations more directly according to the intrinsic statistical properties learned from data. In this sense, it can be viewed as a more flexible continuation of the same feature-space idea: rather than defining anomaly boundaries through a fixed manifold or hypersphere prior, it lets the representation model and perturbation mechanism jointly determine how abnormal features are formed. Representative methods such as SimpleNet Liu et al. (2023) and UniAD You et al. (2022) synthesize anomalies by perturbing extracted features directly in latent space, while SuperSimpleNet Rolih et al. (2025) further constrains perturbations to more specific regions and introduces a refined segmentation head for better localization-oriented behavior. DSR Zavrtanik et al. (2022) takes another route by learning a codebook and replacing masked feature contents through dual-subspace re-projection, thereby generating abnormal representations in a more structured manner. Compared with prior-dependent synthesis, this subcategory is usually more flexible because it does not depend on a hand-specified latent geometry and can adapt more naturally to the learned feature distribution. It is therefore often more effective when the underlying anomaly structure is difficult to characterize with a simple prior. At the same time, its performance depends heavily on the quality of the learned latent space. If the representation is weak or poorly organized, perturbations may produce unrealistic or unin-

formative abnormal features, which in turn can limit downstream utility. Moreover, although data-driven methods are often more adaptive, they still mainly operate in representation space and therefore inherit the broader limitation of weak image-level realism.

Table 5: Practical profile of distribution hypothesis-based IAS.

Family	**Resource burden**			**Practical ceiling**	
	Train	**Infer**	**Memory**	**Control**	**Realism**
Prior-dependent	◐	○	◐	◐	○
Data-driven	◐	○	◐	○	○

○/◐/●= low/medium/high. Train/Infer/Memory denote relative training cost, inference cost, and memory footprint, respectively.

Table 5 summarizes the practical characteristics of distribution hypothesis-based IAS from an engineering perspective. Unlike hand-crafted pipelines, these methods require non-trivial training because anomaly synthesis is built on learned representations, such as encoders, autoencoders, feature extractors, or auxiliary latent components. Their inference overhead nevertheless remains relatively low, since synthesis usually involves feature extraction followed by latent perturbation, projection, or replacement rather than iterative image-space generation, while memory usage is typically dominated by the backbone representation model and, when applicable, lightweight modules such as codebooks or auxiliary heads. Overall, the two subcategories reflect different trade-offs within the same representation-space paradigm. Prior-dependent methods are generally more structured and may provide clearer boundary-aware synthesis behavior, but their reliance on predefined latent geometry can reduce flexibility when anomaly structure is complex. By contrast, data-driven methods are usually more adaptive and make fewer assumptions about feature distribution, although their effectiveness depends more strongly on the quality of learned representations and perturbation design. As a result, distribution hypothesis-based synthesis can be viewed as a lightweight intermediate paradigm: it is more principled than hand-crafted augmentation and often useful for detection- or classification-oriented settings, but its limitations become more apparent when downstream tasks require explicit spatial realism, high-fidelity texture synthesis, or tightly aligned image–mask supervision.

5 GM-based Synthesis

As shown in Fig. 5, GM-based synthesis learns explicit image-space generative priors to produce realistic samples with anomalies. Compared with hand-crafted synthesis and distribution hypothesis-based synthesis, this paradigm operates more directly in pixel space and therefore usually offers a substantially higher realism ceiling. At the same time, the benefits of learned generative priors come with higher training and inference cost, as well as a stronger need to preserve structural consistency between synthesized anomalies and normal backgrounds. The main distinction within this paradigm lies in whether the model synthesizes an image with anomalies from noise, translates a normal image into the abnormal domain, or edits only designated local regions.

Full-image synthesis represents the most direct generative route in this paradigm, because it learns to synthesize samples with anomalies directly from random noise. Typical methods use GANs or diffusion models to approximate the anomaly distribution and then generate visually rich images with anomalies through learned sampling. Early representative work such as Multistage GAN Liu et al. (2019) decouples texture synthesis from background–anomaly fusion to better model both local anomalies and contextual coherence. Con-GAN Du et al. (2022) further addresses data scarcity through shared augmentation and a hypersphere-based objective, while DFMGAN Duan et al. (2023) adopts a two-stage strategy that first learns from normal samples and then fine-tunes on abnormal ones. More recent methods push realism further by incorporating stronger image priors. Defect Spectrum Yang et al. (2025) combines large receptive fields with patch-level refinement to model global structure and local detail jointly, and RealNet Zhang et al. (2024) synthesizes more realistic global anomaly content by perturbing variance during reverse diffusion. Because these methods directly model anomaly appearance, they usually provide the strongest realism and diversity among GM-based subcategories. This makes them valuable for augmentation scenarios that emphasize

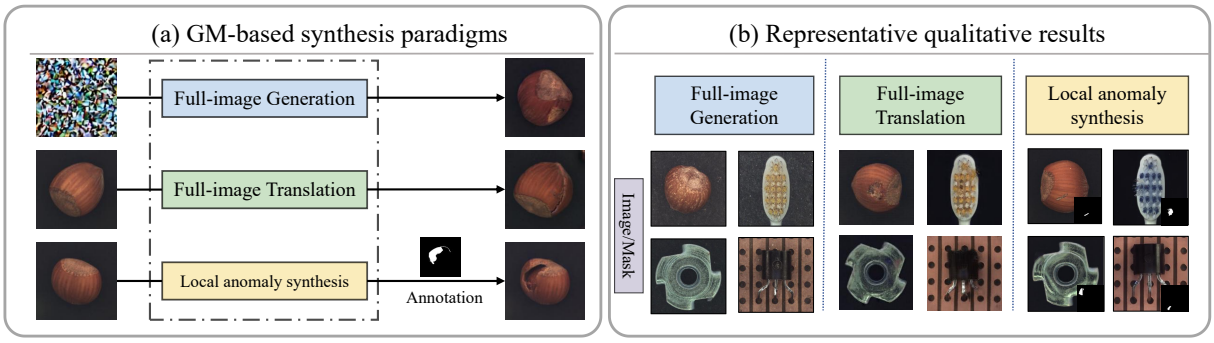

Figure 5: GM-based synthesis. (a) Overview of the GM-based paradigm. It employs deep generative models, such as GANs and diffusion models, for anomaly synthesis, including full-image synthesis, full-image translation, and local anomaly synthesis with region-wise editing. (b) Representative qualitative results adapted from SDGAN Niu et al. (2020), RealNet Zhang et al. (2024) and SARD Wang et al. (2025).

visually rich anomaly content. However, they are also highly dependent on the quantity and diversity of available training data. When real abnormal samples are scarce, the generated images may become unstable, artifact-prone, or insufficiently aligned with real industrial anomaly morphology. In addition, because the entire image is synthesized rather than edited from a given normal sample, preserving fine-grained structural consistency with specific source content can be more difficult.

Full-image translation takes a different route by starting from a normal image and mapping it into the abnormal domain. In this sense, it can be viewed as a more structure-preserving alternative to full-image synthesis: rather than generating the whole image from noise, it modifies an existing image while retaining much of its global layout. Representative translation frameworks such as CycleGAN Zhu et al. (2017) and Pix2Pix Isola et al. (2017) provide the general foundation for this subcategory, and later industrial methods adapt them to anomaly synthesis more explicitly. AttenCGAN Wen et al. (2022) extends cycle-consistent adversarial translation with attention mechanisms, SDGAN Niu et al. (2020) refines translation through additional discriminators for anomaly-oriented sample generation, and Defect-GAN Zhang et al. (2021) explicitly models both anomaly injection and restoration while using spatial distribution maps to preserve normal background appearance. Compared with full-image synthesis, this subcategory often achieves stronger contextual coherence because the generated abnormal image remains anchored to a real normal sample. It is therefore particularly attractive when abundant normal data are available but abnormal data are limited. At the same time, translation-based methods usually offer weaker control over the exact type, location, and extent of synthesized anomalies. Since the transformation is performed at the full-image level, it remains difficult to enforce strict region-level placement or fine-grained mask consistency unless additional conditioning is introduced.

Local anomaly synthesis narrows the generation scope further by editing only selected regions instead of the entire image. Relative to the previous two subcategories, it provides a more targeted synthesis mechanism: the anomaly content is injected or generated within designated local areas, while the remaining background is largely preserved. Representative methods such as MDGAN Wei et al. (2022) construct pseudo-normal backgrounds in abnormal regions to emphasize anomaly distributions, DCDGANc Wei et al. (2023) trains on anomaly-only textures and blends them with different backgrounds for diversified multi-class synthesis, Open-set Fabric Gao et al. (2025) explicitly encodes anomaly type and mask shape to transfer localized fabric anomalies under region constraints, and SARD Wang et al. (2025) performs mask-constrained diffusion synthesis that updates only anomaly regions while preserving the surrounding background for segmentation-aware generation. Because the synthesis region is spatially restricted, this subcategory usually provides the strongest controllability within GM-based IAS and is naturally more compatible with segmentation-oriented downstream tasks. It also alleviates one of the major drawbacks of full-image generation, namely the risk that unrealistic global backgrounds interfere with downstream learning. However, local realism is still not guaranteed. The quality of generated results depends heavily on how well the synthesized anomaly texture blends with surrounding normal content, especially at region boundaries. Moreover, many local editing

pipelines rely on masks or other spatial annotations to define where the anomaly should appear, which increases annotation cost and may limit scalability in practical industrial settings.

Table 6: Practical profile of GM-based IAS.

Family	**Resource burden**			**Practical ceiling**	
	Train	**Infer**	**Memory**	**Control**	**Realism**
Full-image synthesis	◐	◐	◐	◐	●
Full-image translation	◐	◐	◐	○	●
Local anomaly synthesis	◐	●	●	●	●

○/◐/●= low/medium/high. Train/Infer/Memory denote relative training cost, inference cost, and memory footprint, respectively.

Table 6 summarizes the practical characteristics of GM-based IAS from an engineering perspective. In contrast to hand-crafted and representation-space perturbation methods, GM-based synthesis relies on explicit image-space generative priors, and therefore usually involves non-trivial training cost. Inference overhead is also higher, especially when diffusion-based sampling, multi-stage refinement, or auxiliary control modules are introduced, while the runtime footprint is typically larger because these methods often depend on generators, discriminators, encoders, or additional conditioning branches. Overall, the three GM-based subcategories reflect different trade-offs between realism, controllability, and deployment cost. Full-image synthesis offers the greatest generative freedom, but it is also more dependent on abnormal data quality and less constrained by the specific structure of a given source image. Full-image translation improves contextual coherence by transforming normal images toward the anomaly domain, although its anomaly placement and type control usually remain limited without additional conditioning. Local anomaly synthesis provides the strongest spatial controllability and is naturally aligned with segmentation-oriented settings, but its effectiveness depends heavily on region guidance, boundary blending, and annotation quality. As a result, GM-based synthesis can be viewed as a major image-level IAS paradigm that jointly improves realism and controllability, although these gains are achieved at the cost of higher model complexity and computational demand.

6 VLM-based Synthesis

As summarized in Fig. 6, VLM-based synthesis brings multimodal priors into IAS and therefore pushes the field toward more controllable and context-aware anomaly synthesis. Compared with earlier paradigms, its main advantage is that anomaly synthesis is no longer driven only by hand-crafted rules, latent perturbation, or image-space generation alone, but can also be guided by text prompts, spatial conditions, reference examples, or other auxiliary signals. This substantially broadens the controllable interface of IAS and improves semantic richness. At the same time, the main distinction within this paradigm lies in whether anomaly synthesis is completed in a relatively direct single-stage manner or is coupled with additional sequential processes, such as mask generation, alignment correction, or staged refinement.

Single-stage synthesis represents the more direct route in this paradigm, because it relies on pre-trained vision-language or diffusion backbones to synthesize or edit anomalies with minimal additional processing. In most cases, anomaly content is synthesized in one main generation stage under prompt, mask, box, or multimodal guidance, without introducing an explicitly separated refinement pipeline.

Representative methods such as CUT Sun et al. (2024) use training-free diffusion-based guidance to align synthesized anomaly regions with anomaly descriptions, while AnomalyAny Sun et al. (2025) further optimizes cross-attention responses to anomaly-related tokens and refines text semantics through CLIP-guided alignment to synthesize diverse unseen anomalies. Other methods improve controllability or adaptation through lightweight learning. DefectFill Song et al. (2025) uses LoRA fine-tuning to learn anomaly concepts from a few abnormal image–mask pairs and injects them into masked normal images through a single inpainting process, TF-IDG Xu et al. (2025a) combines retrieval and ControlNet-style conditioning to preserve anomaly shape and detail, and AnoGen Gui et al. (2024) learns a compact condition embedding for box-guided inpainting under few-shot supervision.

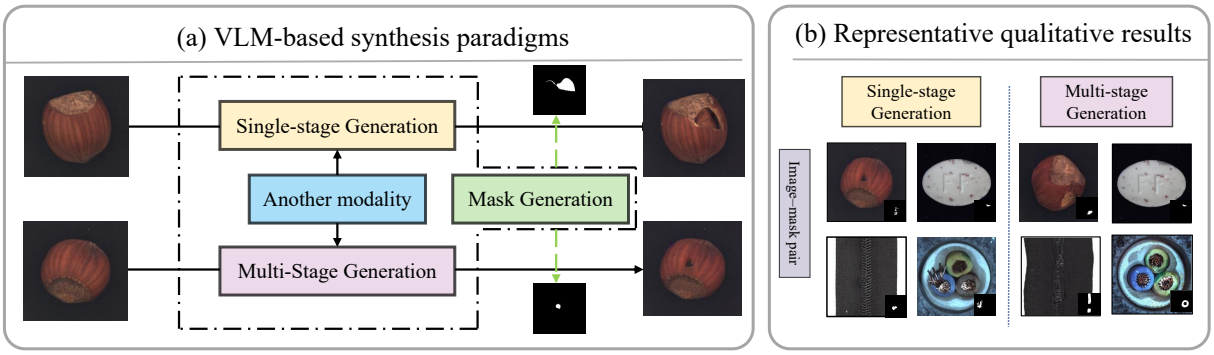

Figure 6: VLM-based synthesis. (a) Overview of the VLM-based paradigm. It leverages large-scale vision–language models with multimodal cues, such as text prompts and spatial conditions, for anomaly synthesis, including single-stage generation and multi-stage generation pipelines with additional mask generation, refinement, or alignment modules. (b) Representative qualitative results adapted from FAST Xu et al. (2025b), AnomalyDiffusion Hu et al. (2024b) and SeaS Dai et al. (2025).

More explicitly multimodal methods such as AnomalyXFusion Hu et al. (2024a), CAGEN Jiang et al. (2024), AnomalyControl He et al. (2024), and FAST Xu et al. (2025b) further enrich this line by introducing dynamic multimodal embeddings, mask-aware diffusion control, cross-modal semantic alignment, or foreground-constrained denoising. Because these methods exploit large-scale pretraining and multimodal cues directly, they usually provide strong realism, efficient deployment, and prompt-level controllability in a relatively concise pipeline. This makes them attractive for rapid prototyping and for scenarios in which users want to specify anomaly semantics without designing complex synthesis systems. However, their synthesis quality still depends on how well spatial conditions are respected during the single-stage process. In practice, synthesized anomaly regions may not always align strictly with target masks or designated positions, which can reduce annotation consistency and limit downstream usefulness for segmentation-oriented tasks.

Multi-stage synthesis builds on the previous idea by coupling anomaly synthesis with additional sequential modules or refinement steps. Instead of completing synthesis in one dominant generation stage, it typically decomposes the process into multiple coordinated stages, such as explicit mask production, concept learning, alignment correction, staged denoising, or coarse-to-fine refinement. In this sense, it can be viewed as a more downstream-oriented extension of VLM-based synthesis: the goal is not only to synthesize visually plausible images with anomalies, but also to improve spatial precision, mask consistency, and practical compatibility with industrial supervision signals.

Representative methods such as DualAnoDiff Jin et al. (2024) use a dual-branch design to model global context and anomaly regions separately while introducing segmentation-assisted mask generation. DefectDiffu Shi et al. (2025) combines textual guidance with dedicated mask generation to improve small-anomaly synthesis and more precise control over anomaly properties. GAA Lu et al. (2025) learns anomaly concept embeddings and then synthesizes semantically aligned anomaly–mask pairs through region-guided mask design, while TDAD Wei et al. (2025) couples multiscale anomaly synthesis with a two-stage diffusion process to suppress over-reconstruction and recover semantic detail.

Table 7: Practical profile of VLM-based IAS.

Family	**Resource burden**			**Practical ceiling**	
	Train	**Infer**	**Memory**	**Control**	**Realism**
Single-stage VLM-based	◐	●	●	◐	●
Multi-stage VLM-based	●	●	●	●	●

○/◐/●= low/medium/high. Train/Infer/Memory denote relative training cost, inference cost, and memory footprint, respectively.

Other pipelines such as BeltDiff Zhuang et al. (2025), SeaS Dai et al. (2025), AnomalyDiffusion Hu et al. (2024b), and AdaBLDM Li et al. (2024) likewise incorporate self-labeling, token-level attribute binding, textual inversion, or online adaptation to progressively improve mask alignment and local realism. Relative to single-stage synthesis, this subcategory is usually more compatible with industrial scenarios that require accurate anomaly placement, stronger image–mask consistency, or paired data generation for downstream training. It often provides better spatial control and more stable annotation quality, which is especially valuable for segmentation-oriented benchmarks and practical pipelines. The trade-off is that the overall system becomes more complex, more resource-intensive, and often harder to deploy. Additional stages also introduce more engineering overhead and may accumulate error if mask quality, intermediate alignment, or refinement objectives are not well designed.

Table 7 summarizes the practical characteristics of VLM-based IAS from an engineering perspective. Unlike earlier paradigms, the computational burden of this family is largely dominated by large pre-trained backbones and multimodal conditioning modules. Even when training is lightweight or partly avoided, runtime memory usage and inference cost often remain high because synthesis still depends on foundation-scale diffusion pipelines, control branches, and cross-modal guidance. Overall, the two VLM-based subcategories reflect different trade-offs between controllability, realism, and pipeline complexity. Single-stage methods usually concentrate complexity within a single synthesis process and are more concise for promptable synthesis or lightweight adaptation of pre-trained models, although they may exhibit weaker spatial precision or mask consistency. By contrast, multi-stage methods distribute complexity across multiple coordinated modules, refinement steps, or auxiliary objectives, and therefore often provide stronger spatial alignment and better downstream compatibility, especially when mask quality and annotation consistency are important. As a result, VLM-based synthesis can be viewed as the most semantically expressive paradigm in IAS, offering strong image-level controllability and realism, while doing so at the cost of higher computational demand and greater pipeline complexity.

7 Future Directions

Table 8: Comparison of IAS subcategories in terms of inputs and outputs.

Paradigm	Subcategory	Abnormal data	Annotation	Prompt	Reference	Level	Output
Hand-crafted Synthesis	Self-contained	–	✓	–	–	Pixel	Abnormal image + mask
	External-dependent	–	✓	–	✓	Pixel	Abnormal image + mask
	Inpainting-based	–	✓	–	–	Pixel	Abnormal image + mask
Distribution hypothesis-based Synthesis	Prior-dependent	–	–	–	–	Latent	Abnormal feature
	Data-driven	–	–	–	–	Latent	Abnormal feature
GM-based Synthesis	Full-image synthesis	✓	○	○	–	Mixed	Abnormal image
	Full-image translation	✓	○	○	–	Mixed	Abnormal image
	Local anomaly synthesis	○	○	○	○	Mixed	Abnormal image
VLM-based Synthesis	Single-stage	○	○	✓	○	Mixed	Abnormal image
	Multi-stage	○	○	✓	○	Mixed	Abnormal image + mask

Symbols indicate usage status: ✓ = required, ○ = optional, and – = not used. Reference includes external texture banks, retrieved exemplars, or similar auxiliary inputs. "Mixed" denotes methods that combine latent-space generation or editing with image-level decoding, control, or refinement.

Tables 8 and 9 jointly provide a cross-paradigm view of IAS subcategories from two complementary perspectives. Table 8 compares their input requirements, synthesis levels, and output forms, whereas Table 9 summarizes their practical characteristics, including computational cost, controllability, primary target tasks, and key trade-offs. Several general patterns can be observed. Hand-crafted and distribution hypothesis-based methods are relatively lightweight, but they usually operate within limited input or representation spaces and therefore remain constrained in realism, output richness, or downstream flexibility. GM-based methods improve image-level realism substantially, yet their controllability and practical cost vary considerably across subcategories. VLM-based methods further broaden the interface of IAS by incorporating prompt-based and multimodal conditions, but such gains are often accompanied by higher computational overhead and more complex pipelines. Taken together, these observations highlight that future IAS research should move be-

Table 9: Comparison of industrial anomaly synthesis subcategories in terms of deployment characteristics.

Paradigm	Subcategory	Cost	Controllability	Downstream tasks	Key trade-off
Hand-crafted Synthesis	Self-contained	Low	Medium	C/D/S	Good source-background consistency Limited anomaly diversity
	External-dependent	Low	Medium	C/D/S	Richer anomaly appearance Higher risk of local inconsistency
	Inpainting-based	Low	Medium	C/D/S	Simple masking-based synthesis Anomalies may look missing or corrupted
Distribution hypothesis-based Synthesis	Prior-dependent	Medium	Medium	C/D	Clear latent-space structure Limited flexibility for complex abnormal shapes
	Data-driven	Medium	Low	C/D	Adaptive feature perturbation Strong reliance on feature quality
GM-based Synthesis	Full-image synthesis	Medium	Low	C/D	High generation freedom Weaker preservation of source-image structure
	Full-image translation	Medium	Medium	C/D	Better domain-level appearance transfer Weak spatial controllability and limited mask reliability
	Local anomaly synthesis	High	High	C/D/S	Strong spatial controllability Dependence on mask quality and boundary blending
VLM-based Synthesis	Single-stage	High	High	C/D/S	Flexible prompt-driven synthesis Weaker spatial precision and mask consistency
	Multi-stage	High	High	C/D/S	Better image–mask alignment Higher pipeline complexity

Note: C, D, and S denote image-level anomaly classification, anomaly detection, and pixel-level segmentation, respectively. The listed tasks indicate common downstream support rather than exclusive applicability. When reliable pixel-level masks or anomaly maps are available, image-level D/C can usually be obtained by score aggregation or thresholding.

yond isolated method design and place greater emphasis on anomaly diversity, controllable synthesis, and multimodal integration.

Improving anomaly diversity remains a central direction. As indicated by Table 8, many existing IAS subcategories either rely on limited anomaly data, operate within relatively restricted synthesis spaces, or produce only feature-level rather than image-level anomaly outputs. Such constraints naturally narrow anomaly coverage and weaken generalization to rare or complex industrial anomalies. This issue is especially evident in methods that depend on scarce real anomaly samples, predefined perturbation rules, or narrowly structured latent assumptions. Future research should therefore develop diversity-oriented synthesis pipelines that can explore underrepresented anomaly modes more effectively. Promising directions include uncertainty-aware generation, self-supervised discovery of rare anomaly patterns, and active selection of insufficiently covered anomaly modes. Coarse-to-fine synthesis strategies, which first establish global anomaly structure and then refine local details, may also help improve both diversity and realism. More broadly, larger and more diverse industrial anomaly datasets will remain important for reducing overfitting to narrow anomaly distributions and for supporting broader anomaly coverage.

Strengthening controllable anomaly synthesis is equally important. Table 9 reveals a clear controllability gap across IAS subcategories. Although local anomaly synthesis and multi-stage VLM-based methods already provide relatively strong control, many other subcategories still offer only limited control over anomaly type, shape, location, or visual intensity. This weakness directly reduces their value for downstream settings that require accurate spatial supervision, precise image–mask alignment, or fine-grained attribute editing. Future IAS research should therefore place greater emphasis on explicit spatial conditioning, attribute-aware generation, and more structured control over anomaly properties. Large-scale generative models, mask-guided editing frameworks, and segmentation-aware synthesis pipelines are especially promising in this respect, because they can better separate anomaly characteristics from normal background content while preserving structural coherence. More generally, controllability should be treated not as an auxiliary extension, but as a central design objective for downstream-oriented IAS.

Promoting multimodal anomaly synthesis in a more systematic manner is another important direction. Table 8 shows that multimodal inputs are still concentrated mainly in VLM-based subcategories, whereas most earlier paradigms remain dominated by image-centric or feature-space-oriented synthesis. This suggests that

the multimodal potential of IAS has not yet been fully explored. In practical industrial scenarios, however, complementary modalities such as text descriptions, infrared measurements, X-ray scans, depth signals, and process metadata can provide richer semantic and structural cues for anomaly synthesis. Future work should therefore investigate cross-modal alignment strategies that connect such heterogeneous signals with visual anomaly generation more effectively. Potential directions include multimodal transformers, contrastive representation learning, retrieval-augmented generation, and multi-source fusion frameworks that jointly model anomaly appearance, spatial layout, and semantic intent. Progress along this line may enable IAS systems to synthesize anomaly samples that are not only more realistic, but also more controllable and better aligned with practical industrial inspection requirements.

8 Conclusion

In this survey, we presented a systematic review of industrial anomaly synthesis (IAS) by examining three central challenges, namely limited anomaly distribution coverage, the difficulty of synthesizing realistic anomaly samples, and the underexplored role of multimodal information. To organize the field in a clearer and more task-oriented manner, we introduced a dedicated taxonomy that groups existing IAS methods into four paradigms: hand-crafted synthesis, distribution hypothesis-based synthesis, GM-based synthesis, and VLM-based synthesis. Based on this taxonomy, we further reviewed representative methods, summarized commonly used datasets and evaluation metrics, and compared the practical characteristics of different IAS subcategories in terms of inputs, outputs, controllability, cost, and primary target tasks.

Overall, IAS has progressed from heuristic and low-cost perturbation toward more realistic, more controllable, and increasingly multimodal synthesis. Hand-crafted and distribution hypothesis-based methods still provide practical value in lightweight or data-limited settings, whereas GM-based and VLM-based methods have pushed IAS closer to image-level realism and downstream-oriented supervision. At the same time, the field still faces clear limitations in anomaly diversity, fine-grained controllability, and multimodal integration. Continued progress in these directions will be important for developing IAS methods that are not only more realistic, but also more useful for industrial detection, segmentation, and related applications. We hope this survey can serve as a structured reference for understanding existing IAS methods and for supporting future research in this area.

References

Toshimichi Aota, Lloyd Teh Tzer Tong, and Takayuki Okatani. Zero-shot versus many-shot: Unsupervised texture anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5564–5572, January 2023.

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 (CVPR)*, 2019.

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, 2022a. doi: 10.1007/s11263-022-01578-9.

Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, pp. 202–213, Setúbal, Portugal, 2022b. SCITEPRESS – Science and Technology Publications. doi: 10.5220/0010865000003124.

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

Jakob Božič, Domen Tabernik, and Danijel Skočaj. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Computers in Industry*, 129:103459, 2021. doi: 10.1016/j.compind.2021.103459.

Yunkang Cao, Xiaohao Xu, Zhaoge Liu, and Weiming Shen. Collaborative discrepancy optimization for reliable image anomaly localization. *IEEE Transactions on Industrial Informatics*, 19(11):10674–10683, 2023.

Yunkang Cao, Xiaohao Xu, Jiangning Zhang, Yuqi Cheng, Xiaonan Huang, Guansong Pang, and Weiming Shen. A survey on visual anomaly detection: Challenge, approach, and prospect. *CoRR*, abs/2401.16402, 2024. URL `https://doi.org/10.48550/arXiv.2401.16402`.

Qiyu Chen, Huiyuan Luo, Han Gao, Chengkan Lv, and Zhengtao Zhang. Progressive boundary guided anomaly synthesis for industrial anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024a.

Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *Computer Vision – ECCV 2024*, pp. 37–54. Springer, 2024b.

Yajun Chen, Yuanyuan Ding, Fan Zhao, Erhu Zhang, Zhangnan Wu, and Linhao Shao. Surface defect detection methods for industrial products: A review. *Applied Sciences*, 11(16):7657, 2021.

Zhewei Dai, Shilei Zeng, Haotian Liu, Xurui Li, Feng Xue, and Yu Zhou. Seas: Few-shot industrial anomaly image generation with separation and sharing fine-tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23135–23144, 2025.

Zongwei Du, Liang Gao, and Xinyu Li. A new contrastive gan with data augmentation for surface defect recognition under limited data. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2022.

Yuxuan Duan, Yan Hong, Li Niu, and Liqing Zhang. Few-shot defect image generation via defect-aware feature manipulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 571–578, 2023.

Fu-You Fan, Lin Zhang, and Yang Dai. Fegan: A feature extraction based approach for gan anomaly detection and localization. *IEEE Access*, 2024. doi: 10.1109/ACCESS.2024.3406438.

Can Gao, Xiujian Chen, Jie Zhou, Jinbao Wang, and Linlin Shen. Open-set fabric defect detection with defect generation and transfer. *IEEE Transactions on Instrumentation and Measurement*, 2025.

Guan Gui, Bin-Bin Gao, Jun Liu, Chengjie Wang, and Yunsheng Wu. Few-shot anomaly-driven generation for anomaly classification and segmentation. In *European Conference on Computer Vision*, pp. 210–226. Springer, 2024.

Shidan He, Lei Liu, and Shen Zhao. Anomalycontrol: Learning cross-modal semantic features for controllable anomaly synthesis. *arXiv preprint arXiv:2412.06510*, 2024.

Lars Heckler-Kram, Jan-Hendrik Neudeck, Ulla Scheler, Rebecca König, and Carsten Steger. The mvtec ad 2 dataset: Advanced scenarios for unsupervised anomaly detection, 2025.

Jie Hu, Yawen Huang, Yilin Lu, Guoyang Xie, Guannan Jiang, Yefeng Zheng, and Zhichao Lu. Anomalyx-fusion: Multi-modal anomaly synthesis with diffusion. *arXiv preprint arXiv:2404.19444*, 2024a.

Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 8526–8534, 2024b.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.

Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Anomaly detection for real-world industrial applications: Benchmarking recent self-supervised and pretrained methods. In *2022 14th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 64–69, 2022. doi: 10.1109/ICUMT57764.2022.9943437.

Bolin Jiang, Yuqiu Xie, Jiawei Li, Naiqi Li, Yong Jiang, and Shu-Tao Xia. Cagen: Controllable anomaly generator using diffusion model. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3110–3114. IEEE, 2024.

Muen Jin and Michael Heizmann. Cutout as augmentation in contrastive learning for detecting burn marks in plastic granules. *Journal of Sensors and Sensor Systems*, 13:63–74, 2024. doi: 10.5194/jsss-13-63-2024.

Ying Jin, Jinlong Peng, Qingdong He, Teng Hu, Hao Chen, Jiafu Wu, Wenbing Zhu, Mingmin Chi, Jun Liu, Yabiao Wang, et al. Dualanodiff: Dual-interrelated diffusion model for few-shot anomaly image generation. *arXiv preprint arXiv:2408.13509*, 2024.

Paul J. Krassnig and Dieter P. Gruber. ISP-AD: A large-scale real-world dataset for advancing industrial anomaly detection with synthetic and real defects, 2025.

Lars Leyendecker, Shobhit Agarwal, Thorben Werner, Maximilian Motz, and Robert H. Schmitt. A study on data augmentation techniques for visual defect detection in manufacturing. In *Bildverarbeitung in der Automation: Ausgewählte Beiträge des Jahreskolloquiums BVAu 2022*, pp. 73–94. Springer Berlin Heidelberg, 2023. doi: 10.1007/978-3-662-66769-9_6.

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*, pp. 9664–9674, 2021.

Hanxi Li, Zhengxun Zhang, Hao Chen, Lin Wu, Bo Li, Deyin Liu, and Mingwen Wang. A novel approach to industrial defect generation through blended latent diffusion model with online adaptation. *arXiv preprint arXiv:2403.10011*, 2024. URL `https://api.semanticscholar.org/CorpusID:268091266`.

Zhenyu Li, Ning Li, Kaitao Jiang, Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Superpixel masking and inpainting for self-supervised anomaly detection. In *Bmvc*, 2020.

Zhuo Li, Yuhao Yan, Xiangheng Wang, Yifei Ge, and Lin Meng. A survey of deep learning for industrial visual anomaly detection. *Artificial Intelligence Review*, 58:279, 2025. doi: 10.1007/s10462-025-11287-7.

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

Juhua Liu, Chaoyue Wang, Hai Su, Bo Du, and Dacheng Tao. Multistage gan for fabric defect detection. *IEEE Transactions on Image Processing*, 29:3388–3400, 2019.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402–20411, 2023.

Yilin Lu, Jianghang Lin, Linhuang Xie, Kai Zhao, Yansong Qu, Shengchuan Zhang, Liujuan Cao, and Rongrong Ji. Generate aligned anomaly: Region-guided few-shot anomaly image-mask pair synthesis for industrial inspection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 11259–11268, 2025.

Wei Luo, Haiming Yao, Wenyong Yu, and Zhengyong Li. Ami-net: Adaptive mask inpainting network for industrial anomaly detection and localization. *IEEE Transactions on Automation Science and Engineering*, 2024.

Shuai Lyu, Dongmei Mo, and Wai keung Wong. Reb: Reducing biases in representation for industrial anomaly detection. *Knowledge-Based Systems*, 290:111563, 2024.

Yu Mao, Ziyang Chen, Ying Liu, Cong Dong, and Kechen Song. A survey on industrial image anomaly detection: methods, benchmarks and rethinks. *Measurement*, 256:118377, 2025. doi: 10.1016/j.measurement.2025.118377.

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)*, pp. 1–6, Kyoto, Japan, June 2021.

Hitoshi Nakanishi, Masahiro Suzuki, and Yutaka Matsuo. Iterative image inpainting with structural similarity mask for anomaly detection. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2021.

Louise Naud and Alexander Lavin. Manifolds for unsupervised visual anomaly detection. *arXiv preprint arXiv:2006.11364*, 2020.

Shuanlong Niu, Bin Li, Xinggang Wang, and Hui Lin. Defect image sample generation with gan for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, 17(3):1611–1622, 2020.

Mingjing Pei, Ningzhong Liu, Bing Zhao, and Han Sun. Self-supervised learning for industrial image anomaly detection by simulating anomalous samples. *International Journal of Computational Intelligence Systems*, 16(1):152, 2023.

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

Blaž Rolih, Matic Fučka, and Danijel Skočaj. Supersimplenet: Unifying unsupervised and supervised learning for fast and reliable surface defect detection. In *International Conference on Pattern Recognition*, pp. 47–65. Springer, 2025.

Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *European Conference on Computer Vision*, pp. 474–489. Springer, 2022.

Qingfeng Shi, Jing Wei, Fei Shen, and Zhengtao Zhang. Few-shot defect image generation based on consistency modeling. In *European Conference on Computer Vision*, pp. 360–376. Springer, 2025.

Woosang Shin, Jonghyeon Lee, Taehan Lee, Sangmoon Lee, and Jong Pil Yun. Anomaly detection using score-based perturbation resilience. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23372–23382, 2023.

Javier Silvestre-Blanes, Teresa Albero-Albero, Ignacio Miralles, Rubén Pérez-Llorens, and Jorge Moreno. A public fabric database for defect detection methods and results. *Autex Research Journal*, 19(4):363–374, 2019. doi: 10.2478/aut-2019-0035.

Jaewoo Song, Daemin Park, Kanghyun Baek, Sangyub Lee, Jooyoung Choi, Eunji Kim, and Sungroh Yoon. Defectfill: Realistic defect generation with inpainting diffusion model for visual inspection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 18718–18727, 2025.

Han Sun, Yunkang Cao, and Olga Fink. Cut: A controllable, universal, and training-free visual anomaly generation framework. *CoRR*, 2024.

Han Sun, Yunkang Cao, Hao Dong, and Olga Fink. Unseen visual anomaly generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 25508–25517, 2025.

Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3):759–776, 2020. doi: 10.1007/s10845-019-01476-x.

Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24511–24520, 2023.

Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22883–22892, June 2024.

Yanshu Wang, Xichen Xu, Xiaoning Lei, and Guoyang Xie. Sard: Segmentation-aware anomaly synthesis via region-constrained diffusion with discriminative mask guidance. In *2025 International Conference on Machine Intelligence and Nature-Inspired Computing (MIND)*, pp. 247–252, 2025. doi: 10.1109/MIND67540.2025.11351901.

Changyun Wei, Hui Han, Yu Xia, and Ze Ji. Tdad: Self-supervised industrial anomaly detection with a two-stage diffusion model. *Computers in Industry*, 164:104192, 2025.

Jing Wei, Zhengtao Zhang, Fei Shen, and Chengkan Lv. Mask-guided generation method for industrial defect images with non-uniform structures. *Machines*, 10(12):1239, 2022.

Jing Wei, Fei Shen, Chengkan Lv, Zhengtao Zhang, Feng Zhang, and Huabin Yang. Diversified and multi-class controllable industrial defect synthesis for data augmentation and transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4444–4452, 2023.

Long Wen, You Wang, and Xinyu Li. A new cycle-consistent adversarial networks with attention mechanism for surface defect classification with small samples. *IEEE Transactions on Industrial Informatics*, 18(12): 8988–8998, 2022.

Matthias Wieler, Tobias Hahn, and Fred A. Hamprecht. Weakly supervised learning for industrial optical inspection. [Dataset], Heidelberg Collaboratory for Image Processing (HCI), Heidelberg University, 2007. URL `https://hci.iwr.uni-heidelberg.de/content/weakly-supervised-learning-industrial-optical-inspection/`. Retrieved from the official dataset page.

Xuan Xia, Xizhou Pan, Nan Li, Xing He, Lin Ma, Xiaoguang Zhang, and Ning Ding. Gan-based anomaly detection: A review. *Neurocomputing*, 493:497–535, 2022.

Ruyi Xu, Yen-Tzu Chiu, Tai-I Chen, Oscar Chew, Yung-Yu Chuang, and Wen-Huang Cheng. Training-free industrial defect generation with diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24214–24223, 2025a.

Xichen Xu, Yanshu Wang, Jinbao Wang, XiaoNing Lei, Guoyang Xie, GUANNAN JIANG, and Zhichao Lu. FAST: Foreground–aware diffusion with accelerated sampling trajectory for segmentation–oriented anomaly synthesis. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL `https://openreview.net/forum?id=qqEfm8tlCM`.

Xichen Xu, Yanshu Wang, Jinbao Wang, Qunyi Zhang, Xiaoning Lei, Guoyang Xie, Guannan Jiang, and Zhichao Lu. Stage: Segmentation-oriented industrial anomaly synthesis via graded diffusion with explicit mask alignment. *arXiv preprint arXiv:2509.06693*, 2025c.

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.

Shuai Yang, Zhifei Chen, Pengguang Chen, Xi Fang, Yixun Liang, Shu Liu, and Yingcong Chen. Defect spectrum: a granular look of large-scale defect datasets with rich semantics. In *European Conference on Computer Vision*, pp. 187–203. Springer, 2025.

Hang Yao, Ming Liu, Haolin Wang, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: Towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *Computer Vision – ECCV 2024*, volume 15129 of *Lecture Notes in Computer Science*, pp. 1–17. Springer, 2024. doi: 10.1007/978-3-031-73209-6_1.

Seungdong Yoa, Seungjun Lee, Chiyoon Kim, and Hyunwoo J. Kim. Self-supervised learning for anomaly detection with dynamic local augmentation. *IEEE Access*, 9:147201–147211, 2021. doi: 10.1109/ACCESS.2021.3124525.

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

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*, pp. 8330–8339, 2021a.

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

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *European conference on computer vision*, pp. 539–554. Springer, 2022.

Gongjie Zhang, Kaiwen Cui, Tzu-Yi Hung, and Shijian Lu. Defect-gan: High-fidelity defect synthesis for automated defect inspection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2524–2534, 2021.

Lin Zhang, Yang Dai, Fuyou Fan, and Chunlin He. Anomaly detection of gan industrial image based on attention feature fusion. *Sensors*, 23(1):355, 2023a. doi: 10.3390/s23010355.

Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16699–16708, 2024.

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

Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, and Hao Zhao. Pad: A dataset and benchmark for pose-agnostic anomaly detection. In *Advances in Neural Information Processing Systems*, volume 36, pp. 44558–44571, 2023.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

Peixian Zhuang, Yuanxiu Cai, Xi Liu, Xianchao Zheng, Fuheng Xiao, and Jiangyun Li. Beltdiff: Diffusion-based self-labeled generation system for conveyor belt damage detection. *Engineering Applications of Artificial Intelligence*, 161:112287, 2025.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Computer Vision – ECCV 2022*, pp. 392–408, Cham, 2022. Springer. doi: 10.1007/978-3-031-20056-4_23.