A Unified Anomaly Synthesis Strategy with Gradient Ascent for Industrial Anomaly Detection and Localization

Qiyu Chen[1,2], Huiyuan Luo[1], Chengkan Lv[1(✉)], and Zhengtao Zhang[1,2,3]

[1] Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences
[3] CASI Vision Technology CO., LTD., Luoyang, China
{chenqiyu2021,huiyuan.luo,chengkan.lv,zhengtao.zhang}@ia.ac.cn

Abstract. Anomaly synthesis strategies can effectively enhance unsupervised anomaly detection. However, existing strategies have limitations in the coverage and controllability of anomaly synthesis, particularly for weak defects that are very similar to normal regions. In this paper, we propose Global and Local Anomaly co-Synthesis Strategy (GLASS), a novel unified framework designed to synthesize a broader coverage of anomalies under the manifold and hypersphere distribution constraints of Global Anomaly Synthesis (GAS) at the feature level and Local Anomaly Synthesis (LAS) at the image level. Our method synthesizes near-in-distribution anomalies in a controllable way using Gaussian noise guided by gradient ascent and truncated projection. GLASS achieves state-of-the-art results on the MVTec AD (detection AUROC of 99.9%), VisA, and MPDD datasets and excels in weak defect detection. The effectiveness and efficiency have been further validated in industrial applications for woven fabric defect detection. The code and dataset are available at: https://github.com/cqylunlun/GLASS.

Keywords: Industrial anomaly detection · Anomaly synthesis · Weak defect detection · Gradient ascent

1 Introduction

Anomaly detection and localization aim to identify and localize abnormal regions by leveraging normal samples. Due to the challenge of collecting sufficient defect samples and the high cost of pixel-level annotations, supervised approaches become impractical in these contexts. Consequently, unsupervised anomaly detection techniques are widely applied in industrial inspection scenarios [2,13,23,38]. Moreover, since the weak defects are anomalies with small areas or low contrast, some abnormal regions may be in close proximity to normal regions.

Existing anomaly detection methods can broadly be classified into three main categories. Reconstruction-based methods [1, 37] detect anomalies by analyzing the residual image before and after reconstruction. Embedding-based methods [8, 15] leverage pre-trained networks to extract and compress features into

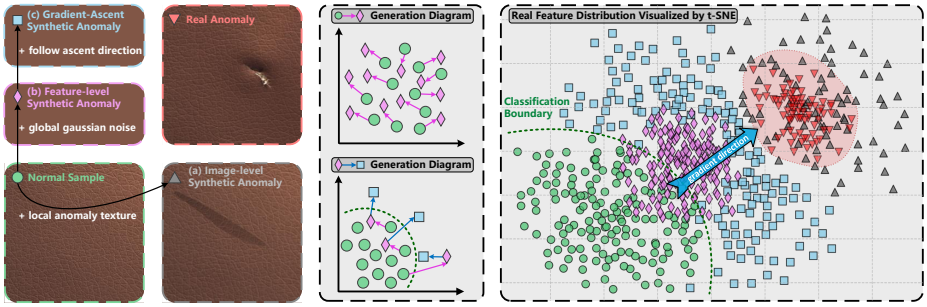

Fig. 1: Process flow and visualization of various anomaly synthesis strategies. (a) Image-level anomaly synthesis strategy (gray triangles) provides detailed textures but lacks diversity. (b) Feature-level anomaly synthesis strategy (pink diamonds) is more efficient but lacks directionality. (c) Our method (blue squares) controls the distribution of synthetic anomalies at image and feature levels by using gradient ascent.

a compact space, distinctly separating anomaly features from normal clusters within the feature space. These two categories are directly trained on original normal samples. However, they fail to resolve the aforementioned issue. Synthesis-based methods [20, 30, 36, 38] typically synthesize anomalies from normal samples, introducing anomaly discrimination information into the detection model for enhanced performance.

A common paradigm is the image-level anomaly synthesis strategy [5, 30, 36], as depicted in Fig. 1(a), which explicitly simulates anomalies at the image level. Although the image-level anomaly synthesis provides detailed anomaly textures, it is considered as lacking diversity and realism. Recent methods [20, 34, 38] are based on the feature-level anomaly synthesis strategy, as illustrated in Fig. 1(b), which implicitly simulates anomalies at the feature level. Due to the smaller size of the feature maps, the feature-level anomaly synthesis is more efficient. However, it also lacks the capability to synthesize anomalies directionally in a controllable way, particularly for near-in-distribution anomalies.

To address the limitations mentioned above, we propose Global and Local Anomaly co-Synthesis Strategy (GLASS), a novel unified framework designed to synthesize a broader coverage of anomalies under the manifold and hypersphere distribution constraints of Global Anomaly Synthesis (GAS) at the feature level and Local Anomaly Synthesis (LAS) at the image level. Specifically, we propose the novel feature-level GAS, as illustrated in Fig. 1(c), which utilizes Gaussian noise guided by gradient ascent and truncated projection. GAS synthesizes anomalies near the normal sample distribution in a controllable way, resulting in a tighter classification boundary that further enhances weak defect detection. The image-level LAS makes improvements by providing a more diverse range of anomaly synthesis. GAS synthesizes weak anomalies around normal points, while LAS synthesizes strong anomalies far from normal points. Theoretically, the near-in-distribution anomalies synthesized by GAS are derived from normal features through relatively small noise and gradient ascents, while the far-from-

distribution anomalies synthesized by LAS are generated by significantly overlaying textures on normal images. Therefore, the rightmost t-SNE visualization of Fig. 1 shows that the anomalies guided by gradient ascent predominantly position themselves close to the appropriate classification boundary. Compared to the anomaly synthesis strategy based on Gaussian noise, our method minimizes the overlap between anomalous and normal samples, reducing the risk of misclassifying normal samples as anomalies.

The main contributions of the proposed GLASS are summarised as follows:

- We propose a unified framework for synthesizing a broader coverage of anomalies in a controllable way at image and feature levels.
- We propose a novel feature-level GAS method that utilizes Gaussian noise guided by gradient ascent to enhance weak defect detection.
- Extensive experiments demonstrate that GLASS outperforms state-of-the-art (SOTA) methods in industrial anomaly detection and localization tasks.

2 Related Work

Reconstruction-based methods such as AutoEncoders [37,40], detect anomalies by analyzing the residual image before and after reconstruction. These methods assume that the model will properly reconstruct normal regions while failing to reconstruct abnormal regions. However, they heavily rely on the quality of reconstructed image and face challenges with the difference analysis method.

Embedding-based methods utilize pre-trained networks to extract features, subsequently compressing normal features into a compact space. As a result, anomaly features are distinctly separated from normal clusters within the feature space. Memory bank methods [2, 12, 23] archive representative normal features and detect anomalies through metric learning. Similarly, one-class classification methods [15, 22, 31] further define explicit classification boundaries, such as hyperplanes [27] or hyperspheres [29]. Normalizing flow [9] methods [11, 16, 35] aim to transform the distribution of normal samples into a standard Gaussian distribution, causing anomalies to exhibit low likelihood. Knowledge distillation methods [3, 8, 25] leverage the distinction in anomaly detection capabilities between teacher and student networks. Despite achieving good performance, these feature embedding methods are only trained on original normal samples, lacking the representation of anomaly samples.

Synthesis-based methods view the synthesis of anomalies as a form of data augmentation from the normal samples. The objective is to introduce anomaly discrimination information and mitigate potential overfitting that may arise from mapping all normal samples to one point. Most existing methods synthesize anomalies at the image level: CutPaste [17] employs a straightforward approach by cutting normal regions and pasting them at random positions; NSA [26] uses Poisson image editing to seamlessly blend blocks of different sizes from various images, synthesizing a series of anomalies that are more similar to natural sub-image irregularities; DRAEM [36] synthesizes anomalies by creating binary masks using Perlin noise and filling them with external textures in the

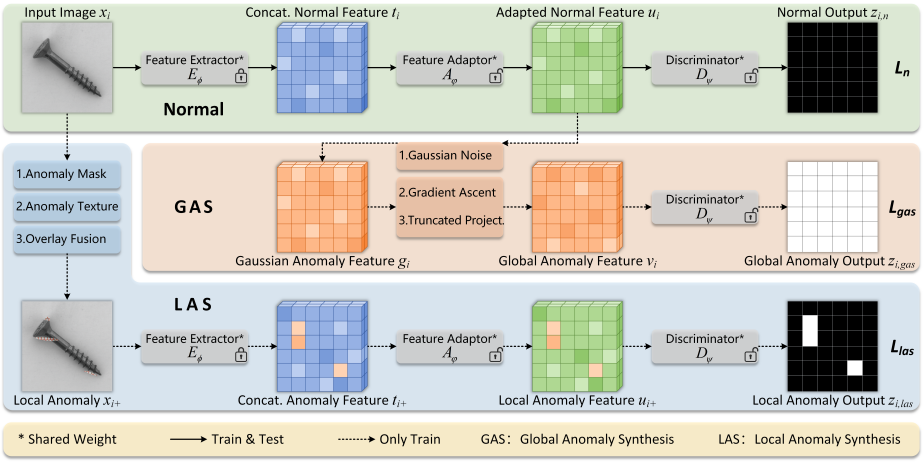

Fig. 2: Schematic of the proposed GLASS. The training phase comprises three branches: (a) Normal branch obtains adapted normal features through a feature extractor and a feature adaptor. (b) GAS branch synthesizes global anomaly features from normal features in three steps based on gradient guidance. (c) LAS branch synthesizes local anomaly images from normal images in three steps based on texture overlay.

normal images. Recently, several methods synthesize anomalies in the feature space: DSR [38] samples in quantified feature space and synthesizes weak defects through the similarity comparison of codebook feature vector; SimpleNet [20] and UniAD [34] synthesize anomalies by adding Gaussian noise to the normal features. Generally, image-level anomaly synthesis provides detailed anomaly textures but lacks diversity, whereas feature-level anomaly synthesis is more efficient but faces challenges with directionality and controllability.

3 Proposed Method

The overall architecture of the proposed GLASS is shown in Fig. 2. During the training stage, GLASS primarily consists of three branches: Normal branch, GAS branch, and LAS branch. Each branch shares three modules: a feature extractor E_ϕ, a feature adaptor A_φ, and a discriminator D_ψ. Normal samples are first processed by the frozen E_ϕ and the trainable A_φ to obtain adapted normal features in Normal branch. Next, global anomaly features are synthesized from the adapted normal features using gradient guidance in GAS branch. Meanwhile, local anomaly images are synthesized by LAS branch through texture overlay, which are then processed by E_ϕ and A_φ to obtain local anomaly features. Finally, the three features from the three branches are jointly fed into the discriminator D_ψ, which is a segmentation network trained end-to-end using three loss functions. During the inference phase, only the framework of normal branch is used to process the test images.

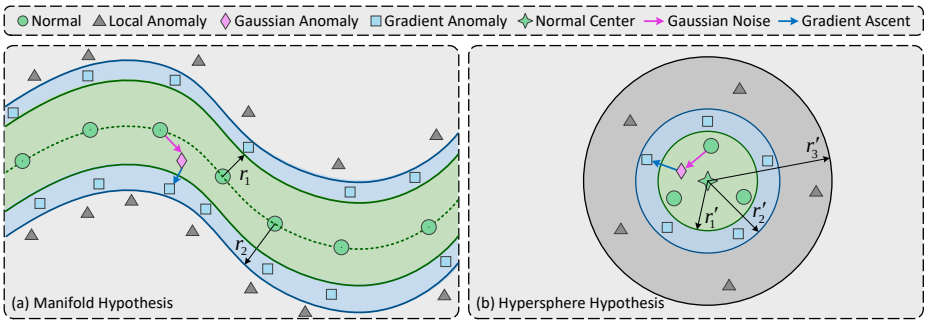

Fig. 3: Schematic illustration of Global Anomaly Synthesis (GAS) under different hypotheses. Assume that r_m or r_h represents the L_2 distance to manifold or hypersphere center, respectively. Green circles ($r_m < r_1$ or $r_h < r'_1$) represent normal features, gray triangles ($r_m > r_2$ or $r'_2 < r_h < r'_3$) represent local anomaly features, pink diamonds represent Gaussian anomaly features obtained by Gaussian noise from normal features, and blue squares ($r_1 < r_m < r_2$ or $r'_1 < r_h < r'_2$) represent global anomaly features obtained by gradient ascent and truncated projection from Gaussian anomaly features.

3.1 Feature Extractor and Feature Adaptor

Similar to [15, 20], we utilize A_φ to mitigate latent domain bias brought by the frozen E_ϕ. The feature map for image $x_i \in X_{\text{train}}$ of level j extracted by the pre-trained backbone ϕ is denoted as $\phi_{i,j} = \phi_j(x_i) \in \mathbb{R}^{H_j \times W_j \times C_j}$. The vector at location (h, w) is represented as $\phi_{i,j}^{h,w} \in \mathbb{R}^{C_j}$. By aggregating the neighborhood features through adaptive average pooling, the locally aware vector $s_{i,j}^{h,w} \in \mathbb{R}^{C_j}$ is derived from the neighborhood features of $\phi_{i,j}^{h,w}$ with neighborhood size p. The set of vectors $s_{i,j}^{h,w}$ constitutes the feature map $s_{i,j}$. By upsampling and merging $s_{i,j}$ from different levels, the concatenated feature map $t_i \in \mathbb{R}^{H_m \times W_m \times C}$ is denoted as $t_i = E_\phi(x_i)$, where the channel size $C = \sum_{j \in J} C_j$. The adapted normal vector $u_i^{h,w}$ is denoted as $u_i^{h,w} = A_\varphi(t_i^{h,w})$, where A_φ employs a single-layer perceptron with the same number of nodes in the input and output layers.

3.2 Feature-level Global Anomaly Synthesis Strategy

Synthesizing anomalies in the feature space [20, 34, 38] has been proven to be an effective method. However, existing methods lack the capability to synthesize anomalies directionally in a controllable way, particularly for near-in-distribution anomalies. To more efficiently synthesize anomalies in feature space, we obtain global anomaly features by adding Gaussian noise to normal features and constraining the synthetic direction of these anomalies using gradient ascent. Here, "global" implies that anomalies are synthesized across all points of the feature map. To avoid the excessive fluctuation of gradient ascent and make the anomaly synthesis more controllable, truncated projection is employed to limit the minimum and maximum range of gradient ascent. The GAS is described as follows:

Distribution Hypothesis. It is posited that all normal feature points conform to either a manifold or a hypersphere distribution hypothesis [21]. The manifold hypothesis assumes that the set of all normal feature points $u_i^{h,w}$, denoted by $U \subseteq \mathbb{R}^C$, satisfies a low-dimensional locally linear manifold distribution [10]. Since manifolds are locally linear and homeomorphic to Euclidean space, the linear combination of low-dimensional embeddings can represent the global nonlinear distribution. Under the manifold hypothesis illustrated in Fig. 3(a), the feature set $N_a = \{\tilde{u}_i^{h,w} \mid \|\tilde{u}_i^{h,w} - u_j^{h,w}\|_2 > r_1, \forall u_j^{h,w} \in U\}$ is considered as anomalous. The hypersphere hypothesis assumes that the set of $u_i^{h,w}$ can be encompassed by a compact hypersphere [24]. Under the hypersphere hypothesis illustrated in Fig. 3(b), the feature set $N_a' = \{\tilde{u}_i^{h,w} \mid \|\tilde{u}_i^{h,w} - c\|_2 > r_1'\}$ is considered as anomalous, where the center of hypersphere is defined as $c = \frac{1}{|U|}\sum_{u_i^{h,w} \in U} u_i^{h,w}$.

Under the manifold and hypersphere hypothesis, the proposed GAS adopts a three-step method involving Gaussian noise, gradient ascent, and truncated projection to synthesize global anomaly features. The first two steps of GAS are the same for manifold and hypersphere hypotheses.

Gaussian Noise. In real-world industrial settings, the distribution of anomalies is unknown. Similar to [20, 34], Gaussian noise is adopted to simulate these diverse anomalies. Specifically, the Gaussian anomaly feature point $g_i^{h,w}$ is obtained by the addition of $u_i^{h,w}$ and noise $\varepsilon_i^{h,w} \sim N(\mu_g, \sigma_g^2)$, denoted as $g_i^{h,w} = u_i^{h,w} + \varepsilon_i^{h,w}$. However, these Gaussian anomaly feature points are synthesized in an undirected way, leading to ineffective training for detection.

Gradient Ascent. The most effective way to synthesize anomalies in feature space is to follow the direction of gradient ascent. Leveraging the previously mentioned Gaussian noise, we integrate gradient information guided by the GAS branch loss L_{gas} in Eq. 6. We normalize the gradient vector and employ a learning rate η for the iterative acquisition of gradient anomaly feature $\tilde{g}_i^{h,w}$ as:

$$\tilde{g}_i^{h,w} = g_i^{h,w} + \eta \frac{\nabla L_{gas}(g_i^{h,w})}{\left\| \nabla L_{gas}(g_i^{h,w}) \right\|} \tag{1}$$

Truncated Projection (Manifold). Although $\tilde{g}_i^{h,w}$ is derived from adding Gaussian noise to normal feature $u_i^{h,w}$ and guided by gradient ascent, there remains a risk of it being either too far from or too close to the normal feature. Therefore, we propose truncated projection to constrain the range of gradient ascent, facilitating controllable anomaly synthesis. The gradient ascent distance is calculated by $\tilde{\varepsilon}_i^{h,w} = \tilde{g}_i^{h,w} - u_i^{h,w}$. To project $\tilde{g}_i^{h,w}$ onto the set $N_p = \{\tilde{g}_i^{h,w} \mid r_1 < \|\tilde{g}_i^{h,w} - u_i^{h,w}\|_2 < r_2\}$ in Fig. 3(a), the truncated distance $\hat{\varepsilon}_i^{h,w}$ is given by:

$$\hat{\varepsilon}_i^{h,w} = \frac{\alpha_i}{\left\| \tilde{\varepsilon}_i^{h,w} \right\|} \tilde{\varepsilon}_i^{h,w}, \text{where } \alpha_i = \begin{cases} r_1 & \left\| \tilde{\varepsilon}_i^{h,w} \right\| < r_1 \\ r_2 & \left\| \tilde{\varepsilon}_i^{h,w} \right\| > r_2 \\ \left\| \tilde{\varepsilon}_i^{h,w} \right\| & \text{otherwise} \end{cases} \tag{2}$$

Algorithm 1 GAS under Manifold Hypothesis

Input: normal feature map u_i, number of batch n_{batch}, number of iteration n_{step}, interval of projection n_{proj}
Output: global anomaly feature map v_i
for batch = 1 to n_{batch} **do**
  Initialize u_i by E_ϕ and A_φ
  Gaussian noise. Add ε_i to $u_i \to g_i$
  for step = 1 to n_{step} **do**
    Gradient ascent.
    (a) Calculate the loss L_{gas} of GAS branch by g_i
    (b) Update $\tilde{g}_i$ according to Eq. 1 with no grad.
    if step is a multiple of n_{proj} **then**
      Truncated projection.
      (c) Get gradient ascent distance $\tilde{\varepsilon}_i = \tilde{g}_i - u_i$
      (d) Constrain the range by Eq. 2 to get truncated distance $\tilde{\varepsilon}_i \to \hat{\varepsilon}_i$
      (e) Get GAS feature $v_i = u_i + \hat{\varepsilon}_i$
    end if
  end for
end for
return v_i

where the truncated coefficient α_i depends on the magnitude of gradient ascent distance $\|\tilde{\varepsilon}_i^{h,w}\|$. The manifold distance r_1 and r_2 are constants, typically $r_2 = 2r_1$. Finally, the global anomaly feature $v_i^{h,w} = u_i^{h,w} + \hat{\varepsilon}_i^{h,w}$ is obtained. GAS algorithm under manifold hypothesis is presented in Alg. 1.

Truncated Projection (Hypersphere). Hypersphere hypothesis further constraints the distribution of gradient anomaly features $\tilde{g}_i^{h,w}$ from GAS and local anomaly features $u_{i+}^{h,w}$ from LAS. Similar to Eq. 2, global anomaly feature $\tilde{v}_i^{h,w}$ is obtained by projecting $\tilde{g}_i^{h,w}$ onto the set $N_p' = \{\tilde{g}_i^{h,w} \mid r_1' < \|\tilde{g}_i^{h,w} - c\|_2 < r_2'\}$. Since $u_{i+}^{h,w}$ is generally further away from the normal feature $u_i^{h,w}$ than $\tilde{v}_i^{h,w}$, it is also projected onto the set $N_p'' = \{u_{i+}^{h,w} \mid r_2' < \|u_{i+}^{h,w} - c\|_2 < r_3'\}$ in Fig. 3(b). This is because $u_{i+}^{h,w}$ is unlikely to merge with $u_i^{h,w}$ after truncated projection, which is a problem that might occur under the manifold hypothesis. To make the normal samples more compact, the lower bound threshold r_1' denotes the radius of hypersphere, which is iteratively updated and empirically set to cover 75% of the normal samples. This prevents synthetic anomalies from being too close to the center. The upper bound threshold is typically set as $r_3' = 2r_2' = 4r_1'$.

Given the complex nonlinear structure of manifold distribution, we posit that a more concentrated intraclass distribution aligns more closely with hypersphere distribution, and vice versa. It is confirmed by the experiments that manifold distribution performs slightly better than hypersphere distribution due to the complex nonlinear structures of most defects. In practice, we analyze the image-level spectrogram to determine the distribution hypothesis of different categories. Details for the choice of hypothesis are provided in Sec. B of the appendix.

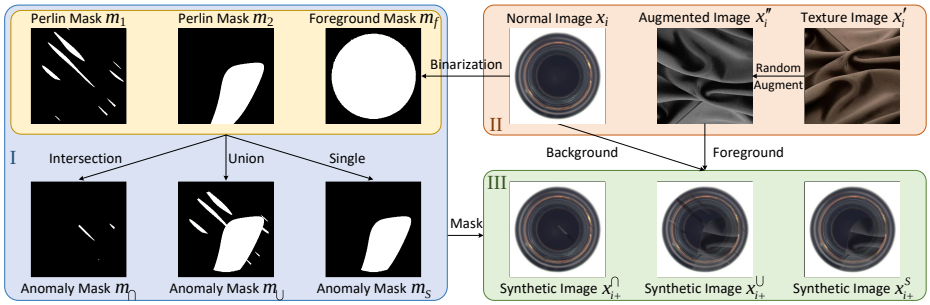

Fig. 4: Flowchart of Local Anomaly Synthesis (LAS) consisting of three steps: Step I: Anomaly Mask, Step II: Anomaly Texture, and Step III: Overlay Fusion.

3.3 Image-level Local Anomaly Synthesis Strategy

Synthesizing anomalies within local regions can provide detailed anomaly textures. Fusing DTD textures with Perlin masks to synthesize anomalies at the image level is a commonly used approach in anomaly detection [32,36,39]. Building on this approach, we propose the image-level LAS to synthesize a more diverse range of anomalies. Fig. 4 presents the flowchart of LAS, detailed as follows:

Anomaly Mask. We first generate two binary masks by Perlin noise, denoted as m_1 and m_2. Since anomalies generally appear on the surface of industrial samples, the foreground mask m_f of normal sample is obtained through binarization inspired by [33]. To increase the diversity of anomalous regions, the intersection and union of m_1 and m_2 is utilized to construct the final mask m_i as:

$$m_i = \begin{cases} (m_1 \wedge m_2) \wedge m_f & 0 \le p_m \le \alpha \\ (m_1 \vee m_2) \wedge m_f & \alpha < p_m \le 2\alpha \\ m_1 \wedge m_f & 2\alpha < p_m \le 1 \end{cases} \tag{3}$$

where random number $p_m \sim U(0,1)$, with α set to $\frac{1}{3}$ in the experiments.

Anomaly Texture. After determining the shape of the anomalous region, we randomly select an image x_i' from the texture dataset DTD [6]. From the set of $K = 9$ image augmentation methods $T = \{T_1, \ldots, T_K\}$, we randomly choose three methods to form $T_R \subset T$, similar to RandAugment [7]. The augmented anomaly texture image is then obtained as $x_i'' = T_R(x_i')$.

Overlay Fusion. To better simulate weak defects while covering more detailed anomalies, we adopt the transparency coefficient $\beta \sim N(\mu_m, \sigma_m^2)$ to modulate the proportion of training set image x_i within the synthetic image under the anomaly mask. The local anomaly image x_{i+} is fused as:

$$x_{i+} = x_i \odot \bar{m}_i + (1-\beta)x_i'' \odot m_i + \beta x_i \odot m_i \tag{4}$$

where $\bar{m}_i$ is derived by inverting the anomaly mask m_i. Subsequently, x_{i+} is processed through Sec. 3.1 to obtain the local anomaly feature map $u_{i+} = A_\varphi(E_\phi(x_{i+}))$, with the corresponding position denoted as $u_{i+}^{h,w}$ at (h,w).

3.4 Discriminator and Training Objectives

Three groups of features are obtained through three branches, respectively, serving as the input for the discriminator D_ψ. It employs a single hidden layer MLP with Sigmoid, directly outputting the anomaly confidence $z_i^{h,w} \in \mathbb{R}$ for each feature point. The training objectives typically consist of three components.

The first term L_n is given by the Binary Cross-Entropy (BCE) loss between the normal feature discrimination $z_{i,n} = D_\psi(u_i)$ and the ground truth of full-size feature map normal:

$$L_n = \sum_{x_i \in X_{train}} f_{BCE}(z_{i,n}, \mathbf{0}) \tag{5}$$

The second term L_{gas} is given by the BCE loss between the global anomaly feature discrimination $z_{i,gas} = D_\psi(v_i)$ and the ground truth of full-size feature map anomaly:

$$L_{gas} = \sum_{x_i \in X_{train}} f_{BCE}(z_{i,gas}, \mathbf{1}) \tag{6}$$

To address the imbalance issue in binary classification of the local anomaly features, the third term L_{las} is given by the Focal loss [18] between the local anomaly feature discrimination $z_{i,las} = D_\psi(u_{i+})$ and the ground truth of anomaly mask m_i:

$$L_{las} = \sum_{x_i \in X_{train}} f_{Focal}(z_{i,las}, m_i) \tag{7}$$

To filter crucial samples such as weak defects, Online Hard Example Mining (OHEM) [28] is applied to L_{las}. The overall loss function is:

$$L = L_n + L_{gas} + f_{ohem}(L_{las}) \tag{8}$$

3.5 Inference and Anomaly Scoring

As depicted in Fig. 2, the inference process is represented by the solid line without GAS and LAS. Input image $x_i \in X_{\text{test}}$ is processed by Sec. 3.1 to obtain $u_i = A_\varphi(E_\phi(x_i))$. Subsequently, D_ψ gives the segmentation result $z_i = D_\psi(u_i)$. By upsampling the interpolation of $z_i \in \mathbb{R}^{H_m \times W_m}$ to the original image size and applying Gaussian smoothing to mitigate noise, the pixel-level anomaly score S_{AL} used for anomaly localization is obtained as:

$$S_{AL} = f_{smooth}(f_{resize}^{H_0,W_0}(z_i)) \tag{9}$$

Additionally, the image-level anomaly score S_{AD} used for anomaly detection is given by the maximum value of all points in z_i.

Table 1: Comparison of GLASS and its variants with different SOTA methods on each category of MVTec AD. ·/· denotes image-level AUROC% and pixel-level AUROC%.

Category	DSR	PatchCore	BGAD	RD++	SimpleNet	GLASS-m	GLASS-h	GLASS-j
Carpet	99.6/96.0	98.6/99.1	99.8/99.4	**100**/99.2	99.7/98.4	99.8/**99.6**	99.2/99.4	99.8/**99.6**
Grid	**100**/**99.6**	97.7/98.8	99.1/99.4	**100**/99.3	99.9/98.5	**100**/99.4	**100**/99.0	**100**/99.4
Leather	99.3/99.5	**100**/99.3	**100**/99.7	**100**/99.5	**100**/99.2	**100**/**99.8**	**100**/**99.8**	**100**/**99.8**
Tile	**100**/98.6	98.8/95.7	**100**/96.7	99.7/96.6	98.7/97.7	**100**/**99.7**	**100**/99.1	**100**/**99.7**
Wood	94.7/91.5	99.1/95.0	99.5/97.0	99.3/95.8	99.5/94.4	**99.9**/**98.8**	99.7/97.6	**99.9**/**98.8**
Texture Avg.	98.7/97.0	98.9/97.6	99.7/98.4	99.8/98.1	99.6/97.6	**99.9**/**99.5**	99.8/99.0	**99.9**/**99.5**
Bottle	99.6/98.8	**100**/98.5	**100**/98.9	**100**/98.8	**100**/98.0	**100**/99.2	**100**/**99.3**	**100**/**99.3**
Cable	95.3/97.7	99.8/98.4	97.9/98.0	99.3/98.4	**100**/97.5	98.2/98.1	99.8/**98.7**	99.8/**98.7**
Capsule	98.3/91.0	98.1/99.0	97.3/99.1	99.0/98.8	97.8/98.9	**99.9**/**99.4**	99.8/99.3	**99.9**/**99.4**
Hazelnut	97.7/99.1	**100**/98.7	99.3/98.5	**100**/99.2	99.8/98.1	**100**/**99.4**	**100**/99.1	**100**/**99.4**
Metal nut	99.1/94.1	**100**/98.3	99.3/97.7	**100**/98.1	**100**/98.8	**100**/**99.4**	**100**/99.1	**100**/**99.4**
Pill	98.9/94.2	96.4/97.8	98.8/98.0	98.4/98.3	98.6/98.6	**99.4**/**99.5**	99.3/99.4	99.3/99.4
Screw	95.9/98.1	98.4/99.5	92.3/99.2	98.9/99.7	98.7/99.2	99.5/**99.5**	**100**/**99.5**	**100**/**99.5**
Toothbrush	**100**/**99.5**	**100**/98.6	86.9/98.7	**100**/99.1	**100**/98.5	**100**/99.3	**100**/99.0	**100**/99.3
Transistor	96.3/80.3	99.9/96.1	99.7/93.9	98.5/94.3	**100**/97.0	99.0/95.5	99.9/**97.6**	99.9/**97.6**
Zipper	98.5/98.4	99.4/98.9	97.8/98.7	98.6/98.8	99.9/98.9	**100**/**99.6**	99.9/99.3	**100**/**99.6**
Object Avg.	98.0/95.1	99.2/98.4	96.9/98.1	99.3/98.3	99.5/98.3	99.6/98.9	**99.9**/99.0	**99.9**/**99.2**
Average	98.2/95.8	99.1/98.1	97.9/98.2	99.4/98.3	99.5/98.1	99.7/99.1	99.8/99.0	**99.9**/**99.3**

4 Experiments

4.1 Datasets

Three widely-used real-world public datasets are employed: MVTec AD [4], VisA [41], and MPDD [14]. Additionally, we construct the woven fabric defect detection (WFDD) dataset under industrial settings with 3860 normal and 241 anomaly samples. To evaluate the ability of GLASS in weak defect detection, we create two test sets based on MVTec AD. MVTec AD-manual (MAD-man) consists of five subsets, each constructed by one of five individuals who selected weak defect samples from every category of MVTec AD under unbiased conditions. Due to the scarcity of weak defect, we also synthesize a weak defect test set named MVTec AD-synthesis (MAD-sys) from five texture categories of MVTec AD. MAD-sys consists of four subsets with different levels of weakness, obtained by adjusting $\beta = \{0.1, 0.3, 0.5, 0.7\}$ in Eq. 4. The WFDD, MAD-man, and MAD-sys datasets are released at this website. Details of these datasets are provided in Sec. A of the appendix.

4.2 Implementation Detail

Experimental Settings. We employ WideResnet50 as the backbone of E_ϕ and merge the features of level2 and level3 for GLASS. The neighborhood size p is set to 3. Input images are resized and center cropped to 288×288. For LAS, transparency coefficient $\beta \sim N(0.5, 0.1^2)$ is truncated within the range $[0.2, 0.8]$. For GAS, Gaussian noise $\varepsilon \sim N(0, 0.015^2)$. GLASS-m is based on the manifold hypothesis where $r_1 = 1, r_2 = 2$ in Eq. 2. GLASS-h is based on the hypersphere hypothesis. GLASS-j is a hybrid strategy derived from judgment that integrates GLASS-h and GLASS-m. The choice between GLASS-h and GLASS-m for each category is determined through the image-level spectrogram analysis method. As

the three variants of GLASS are highly similar, most experiments using GLASS-m by default. We utilize the Adam optimizer to train A_φ and D_ψ with learning rates of 0.0001 and 0.0002, respectively. The training epochs are set to 640 and the batch size is 8. All experiments are implemented on an NVIDIA Tesla A800 GPU and an Intel(R) Xeon(R) Gold 6346 CPU @3.10GHz.

Evaluation Metrics. Area Under the Receiver Operating Characteristic Curve (AUROC) is a commonly used evaluation metric in anomaly detection, we use it to evaluate the discriminative ability of models at image and pixel levels. To provide a more comprehensive evaluation of the anomaly localization ability, we also calculate Per-Region Overlap (PRO) at pixel level.

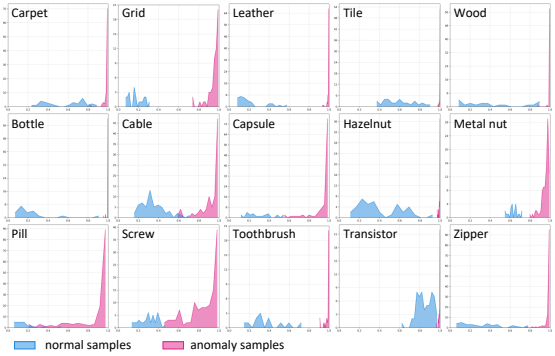

Fig. 5: Anomaly score histograms of GLASS-j on each category of MVTec AD.

4.3 Comparative Experiments on Different Datasets

According to [19], five top SOTA methods across different subfields are employed, including DSR [38], PatchCore [23], BGAD$^{w/o}$ [33], RD++ [30], and SimpleNet [20]. More comparative experiments are provided in Sec. D of the appendix.

Table 2: Comparison of GLASS and its variants with different SOTA methods on four datasets. ·/·/· denotes image-level AUROC%, pixel-level AUROC%, and pixel-level PRO%. The last column provides the throughput measured in img/s.

Method	MVTec AD	VisA	MPDD	WFDD	Avg.	Throughput
DSR [38]	98.2/95.8/91.7	88.0/84.3/61.9	81.0/76.2/58.4	95.1/87.9/78.0	90.6/86.0/72.5	582
PatchCore [23]	99.1/98.1/92.8	94.7/98.5/91.8	93.5/98.9/95.0	96.3/98.1/91.7	95.9/98.4/92.8	31
BGAD [33]	97.9/98.2/96.3	96.4/98.6/92.0	91.8/98.1/93.3	97.1/98.5/88.5	95.8/98.3/92.5	206
RD++ [30]	99.4/98.3/95.0	96.3/98.7/92.2	95.5/98.7/95.6	95.2/98.4/92.9	96.6/98.5/93.9	623
SimpleNet [20]	99.5/98.1/90.0	97.1/98.2/90.7	98.1/98.7/95.7	98.8/98.0/90.6	98.4/98.2/91.8	1306
GLASS-m	99.7/99.1/96.4	**98.8**/98.7/92.5	**99.6/99.4/98.2**	**100/98.9/94.9**	99.5/99.0/95.5	**1327**
GLASS-h	99.8/99.0/95.9	98.2/98.6/90.8	96.7/98.8/96.4	99.0/98.4/88.0	98.4/98.7/92.8	**1327**
GLASS-j	**99.9/99.3/96.8**	**98.8/98.8/92.8**	**99.6/99.4/98.2**	**100/98.9/94.9**	**99.6/99.1/95.7**	**1327**

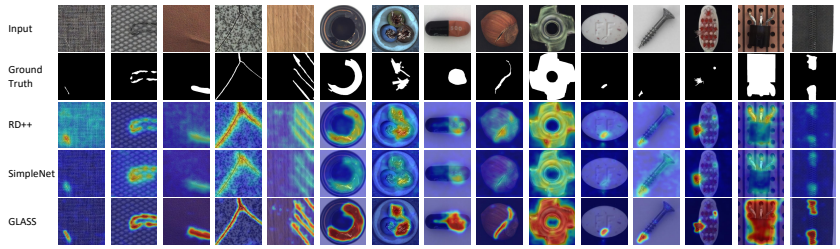

Fig. 6: Qualitative results of GLASS-j and different SOTA methods on MVTec AD.

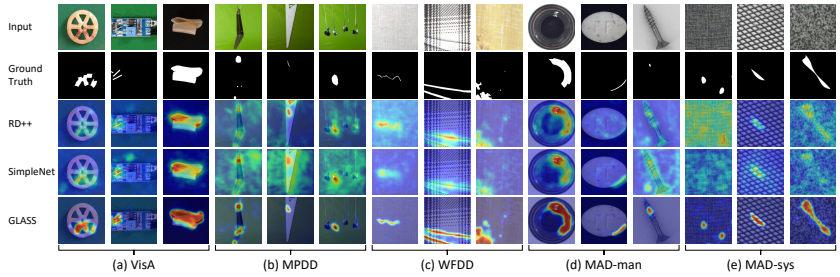

Fig. 7: Qualitative results of GLASS-m and different SOTA methods on datasets.

Anomaly Detection on MVTec AD. As shown in Tab. 1, GLASS-j achieves 100% image-level AUROC on 9 out of 15 categories, further achieving SOTA performance with an average of **99.9%** image-level AUROC and **99.3%** pixel-level AUROC on MVTec AD. Specifically, GLASS-m excels in categories with complex nonlinear structures due to the locally-sensitive manifold distribution, while GLASS-h based on hypersphere distribution is inclined towards categories with concentrated intraclass distribution. The histograms in Fig. 5 show a tiny overlap between normal and abnormal samples, with the anomaly scores for anomalies being markedly high and concentrated. Similarly, Fig. 6 shows that GLASS-j has excellent discriminative ability between normal and abnormal samples.

Table 3: Comparison of GLASS-m with different methods on two weak defect test sets. ·/·/· denotes image-level AUROC%, pixel-level AUROC%, and pixel-level PRO%.

Method	MAD-man	MAD-sys	Avg.
DSR [38]	94.2/96.9/91.3	91.5/90.2/78.3	92.9/93.5/84.8
PatchCore [23]	97.6/98.6/94.7	92.4/92.6/63.9	95.0/95.6/79.3
BGAD [33]	95.9/98.7/96.1	90.4/86.8/64.0	93.2/92.8/80.0
RD++ [30]	98.2/98.9/96.7	83.9/86.3/61.8	91.1/92.6/79.2
SimpleNet [20]	99.3/98.6/94.8	84.6/85.0/57.3	92.0/91.8/76.0
GLASS	**99.6/99.3/97.5**	**95.6/93.3/80.3**	**97.6/96.3/88.9**

Anomaly Detection on Four Datasets. Tab. 2 demonstrates that all three variants of GLASS outperform other SOTA methods across the four datasets

with higher speed. Compared to SimpleNet (based on feature-level anomaly synthesis), GLASS-j increases the average image-level AUROC by 1.2%, pixel-level AUROC by 0.9%, and pixel-level PRO by 3.9%. With a simpler architecture, GLASS achieves superior precision and efficiency on the self-built dataset WFDD collected in industrial settings, further confirming the feasibility of our method. As illustrated in Fig. 7(a-c), GLASS shows outstanding performance in detecting various types of anomalies across different industrial settings.

Anomaly Detection on Weak Defect. Tab. 3 shows the average performance of different methods on MAD-man and MAD-sys, where GLASS surpasses all other methods significantly. Compared to SimpleNet, GLASS achieves improvements of 5.6%, 4.5%, and 12.9% in three metrics, surpassing the level of improvement observed on MVTec AD. Fig. 7(d-e) presents the samples from MAD-man and MAD-sys, showing the outstanding performance of GLASS in weak defect detection. More qualitative results are provided in Sec. E of the appendix.

4.4 Ablation Study

To verify the contribution of different modules, particularly in weak defect detection, we have conducted corresponding ablation experiments mostly on MVTec AD. More ablation studies are provided in Sec. C of the appendix.

Anomaly Synthesis Strategies. We split GAS into three components: Gaussian Noise (GN), Gradient Ascent (GA), and Truncated Projection (TP). As indicated in Tab. 4, GAS (without GA and TP) performs better than LAS on MVTec AD. This indicates that GAS has the advantage of detecting various types of anomalies. However, LAS shows superior performance in weaker defects on MAD-sys, revealing its advantage in detecting local anomalies. The cooperative training of LAS and GAS achieves an obvious improvement, showing their complementarity to synthesize a broader coverage of anomalies.

Table 4: Performance of GLASS-m on MVTec AD and two weak defect test sets under different anomaly synthesis strategies. $\cdot/\cdot/\cdot$ denotes image-level AUROC%, pixel-level AUROC%, and pixel-level PRO%.

LAS	GAS			MVTec AD	MAD-man	MAD-sys
	GN	GA	TP			
✓				98.2/95.4/88.0	97.4/97.1/94.0	94.4/92.0/80.0
	✓			99.4/98.1/91.8	98.4/98.3/95.1	84.1/85.4/60.6
✓	✓			99.5/98.9/94.7	98.7/99.1/96.8	94.6/92.2/77.7
✓	✓	✓		99.6/99.0/95.9	99.0/99.2/97.1	95.0/92.8/79.5
✓	✓	✓	✓	**99.7/99.1/96.4**	**99.6/99.3/97.5**	**95.6/93.3/80.3**

Components in GAS. Tab. 4 explicitly shows that the three evaluation metrics improve successively by adding GN, GA, and TP. As LAS and GAS (without GA and TP) have already achieved 99.5% image-level AUROC on MVTec AD, the introduction of GA and TP offers relative improvement. However, their improvements are more significant on MAD-man and MAD-sys, indicating that GA and TP are particularly effective in detecting weak defects.

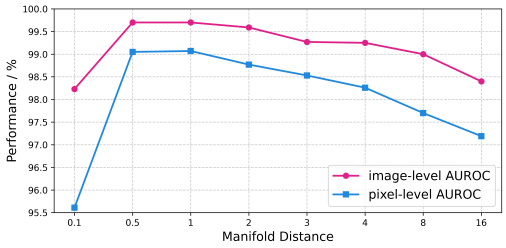

Fig. 8: Performance of GLASS-m on MVTec AD under different manifold distance r_1.

Table 5: Performance of GLASS-m on MVTec AD under different backbone settings. ·/·/· same as above.

Backbone	level1	level2	level3	MVTec AD
WideResNet50	✓			96.7/96.5/90.7
		✓		99.2/97.6/94.6
			✓	**99.7**/98.9/94.6
		✓	✓	**99.7**/**99.1**/**96.4**
	✓	✓		98.4/97.5/92.9
	✓		✓	99.0/98.3/92.3
	✓	✓	✓	99.2/98.6/94.1
ResNet18		✓	✓	99.1/98.0/94.7
ResNet50		✓	✓	99.6/98.9/95.4
ResNet101		✓	✓	99.6/99.0/95.3

Dependency on Backbone. As shown in Tab. 5, merging the features output by level2 and level3 of WideResNet50 achieves the best performance. We have chosen it as the default setting. Meanwhile, our method does not depend on a specific backbone. GLASS can maintain its good performance between several ResNets with different number of parameters on MVTec AD.

Feature Adaptor. As introduced in Sec. 3.1, we utilize the feature adapter A_φ to mitigate latent domain bias brought by the pre-trained backbone of feature extractor E_ϕ. We have conducted experiments using GLASS-m with and without A_φ on MVTec AD. As a result, the absence of A_φ in each branch leads to a decline of 0.1% in pixel-level AUROC and 0.5% in pixel-level PRO.

Manifold Distance. We have introduced the manifold distance r_1 for truncated projection of gradient ascent in Sec. 3.2. It represents the relaxation tolerability for normal feature distribution which should not be too large (overfitting) or too small (underfitting), facilitating controllable anomaly synthesis. As the pre-trained features have already been standardized, the magnitude of gradient ascent distance $\|\tilde{\varepsilon}_i^{h,w}\|$ mostly distributes around 1. Fig. 8 proves that the optimal range of r_1 is $[0.5, 1]$. Therefore, we have chosen $r_1 = 1$ by default.

5 Conclusion

In this paper, we propose a novel unified framework GLASS through the cooperative training of GAS and LAS for synthesizing a broader coverage of anomalies in a controllable way under manifold and hypersphere hypothesis. Specifically, we propose GAS based on gradient ascent and truncated projection. GAS has the capacity for quantitative synthesis of weak defects, solving the problem of random synthetic direction in Gaussian noise. LAS makes improvements by providing a more diverse range of anomaly synthesis. GLASS achieves SOTA results with faster detection speed on four anomaly detection datasets in various industrial settings and shows superior performance in weak defect detection. However, our main focus is localizing the structural anomalies in industrial scenarios. We have not extensively explored the logical anomalies. In the future, we will investigate the application of GLASS in logical anomaly detection and plan to implement anomaly synthesis without relying on auxiliary texture datasets.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 62303458, Grant 62303461 and Grant U21A20482. This work is also supported by the Beijing Municipal Natural Science Foundation of China under Grant L243018. In addition, we would like to express our gratitude to WEIQIAO Textile for collecting the original images used in the WFDD dataset.

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Asia Conference on Computer Vision. pp. 622–637. Springer (2019). https://doi.org/10.1007/978-3-030-20893-6_39
2. Bae, J., Lee, J.H., Kim, S.: Pni: Industrial anomaly detection using position and neighborhood information. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6373–6383 (2023)
3. Batzner, K., Heckler, L., König, R.: Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 128–138 (2024)
4. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9592–9600 (2019)
5. Cao, Y., Xu, X., Liu, Z., Shen, W.: Collaborative discrepancy optimization for reliable image anomaly localization. IEEE Transactions on Industrial Informatics **19**(11), 10674–10683 (2023)
6. Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., Vedaldi, A.: Describing textures in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3606–3613 (2014)
7. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops. pp. 702–703 (2020)
8. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)
9. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real nvp. In: International Conference on Learning Representations (2017)
10. Goyal, S., Raghunathan, A., Jain, M., Simhadri, H.V., Jain, P.: Drocc: Deep robust one-class classification. In: International Conference on Machine Learning. vol. 119, pp. 3711–3721. PMLR (2020)
11. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 98–107 (2022)
12. Hou, J., Zhang, Y., Zhong, Q., Xie, D., Pu, S., Zhou, H.: Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8791–8800 (2021)

13. Hyun, J., Kim, S., Jeon, G., Kim, S.H., Bae, K., Kang, B.J.: Reconpatch: Contrastive patch representation learning for industrial anomaly detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2052–2061 (2024)
14. Jezek, S., Jonak, M., Burget, R., Dvorak, P., Skotak, M.: Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In: International congress on ultra modern telecommunications and control systems and workshops (ICUMT). pp. 66–71. IEEE (2021)
15. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access **10**, 78446–78454 (2022)
16. Lei, J., Hu, X., Wang, Y., Liu, D.: Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14143–14152 (2023)
17. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: 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)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
19. Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F., Jin, Y.: Deep industrial image anomaly detection: A survey. Machine Intelligence Research **21**(1), 104–135 (2024). https://doi.org/10.1007/s11633-023-1459-z
20. Liu, Z., Zhou, Y., Xu, Y., Wang, Z.: 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)
21. Pless, R., Souvenir, R.: A survey of manifold learning for images. IPSJ Transactions on Computer Vision and Applications **1**, 83–94 (2009)
22. Reiss, T., Cohen, N., Bergman, L., Hoshen, Y.: Panda: Adapting pretrained features for anomaly detection and segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2806–2814 (2021)
23. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2022)
24. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International Conference on Machine Learning. pp. 4393–4402. PMLR (2018)
25. Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M.H., Rabiee, H.R.: Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14902–14912 (2021)
26. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural synthetic anomalies for self-supervised anomaly detection and localization. In: European Conference on Computer Vision. pp. 474–489. Springer (2022). https://doi.org/10.1007/978-3-031-19821-2_27
27. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural computation **13**(7), 1443–1471 (2001)
28. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 761–769 (2016)

29. Tax, D.M., Duin, R.P.: Support vector data description. Machine Learning **54**, 45–66 (2004)
30. Tien, T.D., Nguyen, A.T., Tran, N.H., Huy, T.D., Duong, S., Nguyen, C.D.T., Truong, S.Q.: Revisiting reverse distillation for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24511–24520 (2023)
31. Xiao, F., Sun, R., Fan, J.: Restricted generative projection for one-class classification and anomaly detection. arXiv preprint arXiv:2307.04097 (2023)
32. Yang, M., Wu, P., Feng, H.: Memseg: A semi-supervised method for image surface defect detection using differences and commonalities. Engineering Applications of Artificial Intelligence **119**, 105835 (2023)
33. Yao, X., Li, R., Zhang, J., Sun, J., Zhang, C.: Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24490–24499 (2023)
34. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. Advances in Neural Information Processing Systems **35**, 4571–4584 (2022)
35. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. arXiv preprint arXiv:2111.07677 (2021)
36. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8330–8339 (2021)
37. Zavrtanik, V., Kristan, M., Skočaj, D.: Reconstruction by inpainting for visual anomaly detection. Pattern Recognition **112**, 107706 (2021)
38. Zavrtanik, V., Kristan, M., Skočaj, D.: Dsr–a dual subspace re-projection network for surface anomaly detection. In: European Conference on Computer Vision. pp. 539–554. Springer (2022). https://doi.org/10.1007/978-3-031-19821-2_31
39. Zhang, X., Li, S., Li, X., Huang, P., Shan, J., Chen, T.: 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 (2023)
40. Zhou, Y.: Rethinking reconstruction autoencoder-based out-of-distribution detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7379–7387 (2022)
41. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: European Conference on Computer Vision. pp. 392–408. Springer (2022). https://doi.org/10.1007/978-3-031-20056-4_23