RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection

Ximiao Zhang[1] Min Xu[1*] Xiuzhuang Zhou[2]
[1]College of Information and Engineering, Capital Normal University
[2]School of Artificial Intelligence, Beijing University of Posts and Telecommunications
{2211002048, xumin}@cnu.edu.cn[1], xiuzhuang.zhou@bupt.edu.cn[2]

Abstract

Self-supervised feature reconstruction methods have shown promising advances in industrial image anomaly detection and localization. Despite this progress, these methods still face challenges in synthesizing realistic and diverse anomaly samples, as well as addressing the feature redundancy and pre-training bias of pre-trained feature. In this work, we introduce RealNet, a feature reconstruction network with realistic synthetic anomaly and adaptive feature selection. It is incorporated with three key innovations: First, we propose Strength-controllable Diffusion Anomaly Synthesis (SDAS), a diffusion process-based synthesis strategy capable of generating samples with varying anomaly strengths that mimic the distribution of real anomalous samples. Second, we develop Anomaly-aware Features Selection (AFS), a method for selecting representative and discriminative pre-trained feature subsets to improve anomaly detection performance while controlling computational costs. Third, we introduce Reconstruction Residuals Selection (RRS), a strategy that adaptively selects discriminative residuals for comprehensive identification of anomalous regions across multiple levels of granularity. We assess RealNet on four benchmark datasets, and our results demonstrate significant improvements in both Image AUROC and Pixel AUROC compared to the current state-of-the-art methods. The code, data, and models are available at https://github.com/cnulab/RealNet.

Figure 1. SDAS generates anomaly images using only normal images. The example images are sourced from the MVTec-AD dataset [3].

1. Introduction

Image anomaly detection is a critical task in industrial production, with wide-ranging applications in quality control and safety monitoring. While self-supervised methods [18, 29, 42, 44, 46] have gained attention for training models using synthetic anomalies, they still face challenges in synthesizing realistic and diverse anomaly images, especially generating complex structural anomalies and unseen anomaly categories. Due to the lack of available anomaly images and prior knowledge about anomaly categories, existing methods rely on carefully crafted data augmentation strategies [18, 29] or external data [42] for anomaly synthesis, leading to significant distribution discrepancy between synthetic anomalies and real anomalies, thereby limiting the generalization ability of anomaly detection models to real-world applications. To address these issues, we introduce Strength-controllable Diffusion Anomaly Synthesis (SDAS), a novel synthesis strategy that generates diverse samples more closely aligned with natural distributions, and offers flexibility in controlling anomaly strength. SDAS employs DDPM [15] to model the distribution of normal samples and introduces perturbation terms during the sampling process to generate samples in low probability density regions. These samples simulate various natural anomaly patterns, such as aging, structural changes, abnormal textures, and color changes, as shown in Fig. 1.

Parallel to this, feature reconstruction-based anomaly detection [8, 30, 38, 43, 46] is another promising research direction, which reconstructs the features of anomalous images as those of normal images and conducts anomaly detection and localization by reconstruction residuals. They have attracted considerable attention due to the simple

*Corresponding author.

paradigm. However, due to the high computational demands of feature reconstruction and the lack of effective feature selection strategies, existing methods either employ small-scale pre-trained CNNs [30, 38, 43] for anomaly detection or handpick layer-specific features from pre-trained network [8, 46] for reconstruction. The latest work [14] highlights the importance of feature selection, indicating that existing anomaly detection methods [27, 40] are sensitive to feature selection. The optimal pre-trained feature subset for anomaly detection varies across different categories. Therefore, devising a unified feature selection approach has become a pressing need for advancing anomaly detection. In this paper, we propose RealNet, a feature reconstruction framework that incorporates Anomaly-aware Features Selection (AFS) and Reconstruction Residuals Selection (RRS). RealNet fully exploits the discriminative capabilities of large-scale pre-trained CNNs while reducing feature redundancy and pre-training bias, enhancing anomaly detection performance while effectively controlling computational demands. For different categories, RealNet selects different pre-trained feature subsets for anomaly detection, ensuring optimal anomaly detection performance while flexibly controlling the model size. Furthermore, RealNet effectively reduces missed detections by adaptively discarding reconstruction residuals lacking anomalous information, and significantly improves the recall of anomalous regions. In summary, our contributions are fourfold:

- We propose RealNet, a feature reconstruction network that effectively leverages multi-scale pre-trained features for anomaly detection by adaptively selecting pre-trained features and reconstruction residuals. RealNet achieves state-of-the-art performance while addressing the computational cost limitations suffered by previous methods.
- We introduce Strength-controllable Diffusion Anomaly Synthesis (SDAS), a novel anomaly synthesis strategy that generates realistic and diverse anomalous samples closely aligned with natural distributions.
- We evaluate RealNet on four datasets (MVTec-AD [3], MPDD [16], BTAD [22], and VisA [47]), surpassing existing state-of-the-art methods using the same set of network architectures and hyperparameters across datasets.
- We provide the Synthetic Industrial Anomaly Dataset (SIA). SIA is generated by SDAS and consists of a total of 360,000 anomalous images from 36 categories of industrial products. SIA can be conveniently utilized for anomaly synthesis to facilitate self-supervised anomaly detection methods.

2. Related work

Unsupervised anomaly detection and localization approaches use only normal images for model training, without any anomalous data. These methods can be roughly classified into four main categories: reconstruction-based methods [1, 2], self-supervised learning-based methods [18, 42], deep feature embedding-based methods [7, 27], and one-class classification-based methods [20, 37]. In this paper, we focus on the reconstruction-based and self-supervised learning-based methods, which are of particular relevance to our proposed RealNet framework.

Reconstruction-based methods follow a relatively consistent paradigm, which entails training a reconstruction model on normal images. The inability to effectively reconstruct anomalous regions in input images facilitates anomaly detection and localization through comparison of the original and reconstructed images. In this context, a variety of reconstruction techniques are explored, such as Autoencoder [2, 39], GAN [1, 28], Transformer [22, 25], and Diffusion model [21, 33, 45]. However, managing the reconstruction capability of the network remains challenging. In cases of complex image structures or textures, the network may produce a simplistic copy instead of selective reconstruction. Furthermore, inherent stylistic discrepancies between original and reconstructed images can lead to false positives or undetected anomalies.

Recent studies, as exemplified by [8, 30, 38, 43], have been primarily focused on anomaly detection through the reconstruction of pre-trained image features. In contrast to image-level reconstruction, multi-scale features pre-trained on ImageNet [9] demonstrate enhanced discriminative abilities to detect anomalies across a wide range of scales and diverse image patterns. However, due to the inherent feature redundancy in high-dimensional features and the pre-training bias introduced by classification tasks, the anomaly detection capability of large-scale pre-trained networks has not been fully utilized. Recent studies [30, 38, 43] use small-scale pre-trained networks to ensure controllable reconstruction costs, and other works [27, 32, 46] manually select partial layer features from pre-trained networks for anomaly detection. However, the optimal feature subset for anomaly detection varies across different categories [14], thus, these manually selecting methods often prove to be dataset-specific and suboptimal, resulting in a significant performance drop. Different from previous solutions, our RealNet presents a novel combination of efficient feature selection strategies and an optimized reconstruction process, effectively enhancing anomaly detection performance while maintaining computational efficiency.

Self-supervised learning-based methods aim to bypass the need for labels of anomalous images by setting a suitable proxy task. Notable works in this domain include CutPaste [18], which generates anomalies by transplanting image patches from one location to another, albeit with suboptimal continuity in the anomalous regions. NSA [29] uses Poisson image editing [23] for seamless image pasting to synthesize more natural anomaly regions. DRAEM [42] leverages the texture dataset DTD [5] to

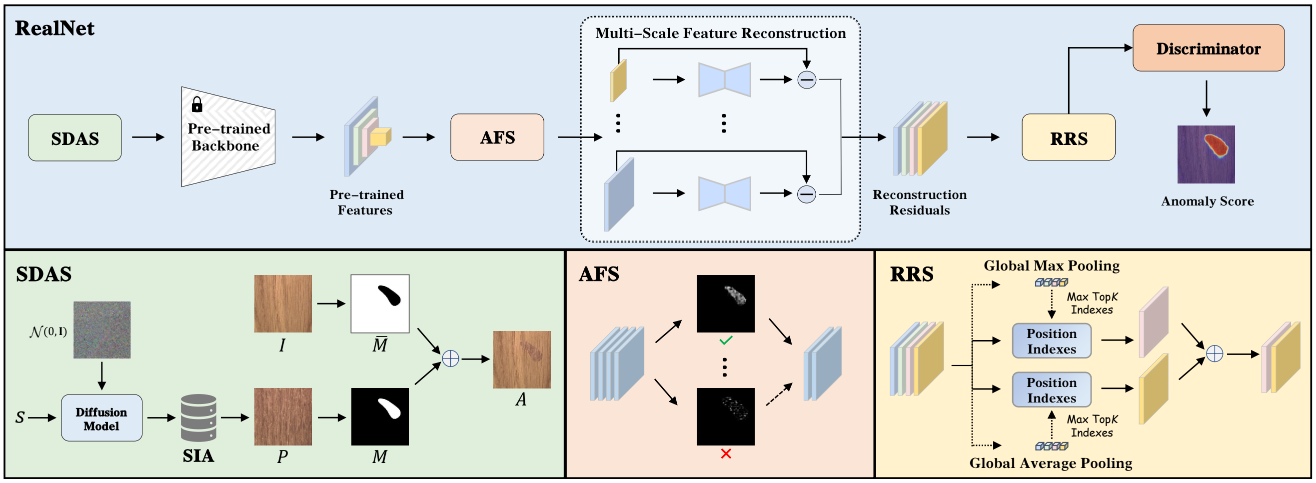

Figure 2. The pipeline of our RealNet consists of three core components: Strength-controllable Diffusion Anomaly Synthesis (SDAS), Anomaly-aware Features Selection (AFS), and Reconstruction Residuals Selection (RRS). 1) SDAS enables the synthesis of diverse, near-natural distribution anomalous images. 2) AFS refines features extracted by large-scale pre-trained CNN for dimensionality reduction. Refined features are reconstructed into corresponding normal image features by a set of reconstruction networks. 3) RRS selects reconstruction residuals most likely to identify anomalies, which are then fed into a discriminator for anomaly detection and localization.

synthesize various texture anomalies and achieve advanced self-supervised anomaly detection performance, however, it falls short when faced with specific structural anomalies, such as partial missing or misplaced elements.

The performance of self-supervised anomaly detection methods hinges on how closely the proxy task aligns with the real anomaly detection task. Anomaly synthesis, as a fundamental study in anomaly detection, has not yet received widespread exploration. Recent work [11] use StyleGAN2 [17] for image editing to generate anomalous images. However, the proposed method relies on real anomalous images and cannot generate unseen anomaly types. In contrast, SDAS operates in the probability space, free from constraints imposed by data augmentation rules or existing data, enabling effective control over anomaly strengths and the generation of realistic and diverse anomaly images using only normal images.

3. Method

In this section, we will introduce our proposed feature reconstruction framework, RealNet, which consists of three key components: Strength-controllable Diffusion Anomaly Synthesis (SDAS), Anomaly-aware Features Selection (AFS), and Reconstruction Residuals Selection (RRS). The pipeline of RealNet is illustrated in Fig. 2.

3.1. Strength-controllable Diffusion Anomaly Synthesis

Denoising Diffusion Probabilistic Models (DDPM) [15] employ a forward diffusion process to incrementally add noise $\mathcal{N}(0, \mathbf{I})$ to the original data distribution $q(x_0)$. At time t, the conditional probability distribution of the noisy data x_t is $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I})$, where $\{\beta_t\}_{t=1}^T$ is a fixed variance schedule, and $\{x_t\}_{t=1}^T$ are the latent variables. The diffusion process is defined as a Markov chain, with joint probability distribution $q(x_{1:T}|x_0) = \prod_{t=1}^T q(x_t|x_{t-1})$. Following the sum rule of Gaussian random variables, the conditional probability distribution of x_t at time t is $q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I})$, where $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$.

The reverse process is described as another Markov chain, where the mean and variance of the reverse process are parameterized by θ, *i.e.*, $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$. There are various ways to model $\mu_\theta(x_t, t)$; typically, neural networks $\epsilon_\theta(x_t, t)$ are used to model the noise ϵ in the diffusion process, resulting in $\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t))$. In the training phase, our goal is to minimize the variational upper bound of the negative log-likelihood, which leads to the simplified objective:

$$\mathcal{L}_{simple} = \mathbb{E}_{t,x_0,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, t)\|^2] \tag{1}$$

To generate realistic anomalous images, we first train a diffusion model to learn the distribution of normal images using Eq. (1). In reverse diffusion process characterized by $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$, x_{t-1} is the normal image obtained at time $t-1$. Due to the anomalous images being located in low-density regions near the normal images, we introduce an additional perturbation $s\Sigma$ to sample anomalous images, yielding $p(x'_{t-1}|x_{t-1}) = \mathcal{N}(x'_{t-1}; x_{t-1}, s\Sigma)$, where Σ is the additional introduced

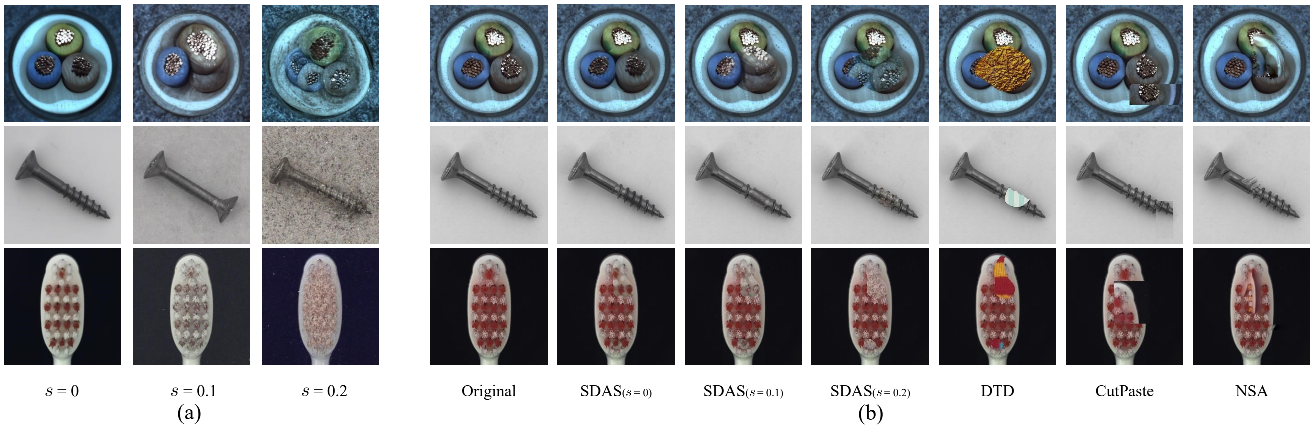

Figure 3. Anomaly image examples generated with different synthesis methods. (a) Examples generated using SDAS with different anomaly strengths s. (b) Examples featuring local anomaly regions generated by various anomaly synthesis methods.

variance, scalar s controls the anomaly strength ($s \geq 0$), and x'_{t-1} is the anomalous image obtained at time $t-1$. To simplify the anomalous synthesis process, we set $\Sigma = \Sigma_\theta(x_t, t)$, by which the conditional probability distribution of anomalous images x'_{t-1} can be written as follows:

$$p_\theta(x'_{t-1}|x_t) = \mathcal{N}(x'_{t-1}; \mu_\theta(x_t, t), (1+s)\Sigma_\theta(x_t, t)) \quad (2)$$

To ensure that the generated anomalous images are close to the distribution of normal images, we set $s \to 0$, resulting in $x'_{t-1} \approx x_{t-1}$; then we use x'_{t-1} for the next time step of the reverse diffusion process. The final form is $p_\theta(x'_{t-1}|x'_t) = \mathcal{N}(x'_{t-1}; \mu_\theta(x'_t, t), (1+s)\Sigma_\theta(x'_t, t))$. We term this process Strength-controllable Diffusion Anomaly Synthesis (SDAS), detailed in Algorithm 1. Specifically, SDAS will generate normal images if s is set to 0.

To incorporate these anomalous images during training of anomaly detection model, we follow the approach presented in [42], utilizing a Perlin noise generator [24] to capture various anomalous shapes and binarize them into an anomaly mask M. We denote the normal image as I, the anomalous image generated by SDAS as P, and the image with local anomalies synthesized by image blending as A:

$$A = \overline{M} \odot I + (1-\delta)(M \odot I) + \delta(M \odot P) \quad (3)$$

Algorithm 1 Strength-controllable Diffusion Anomaly Synthesis (SDAS)

Input: diffusion model $(\mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$
anomaly strength s
$x_T \sim \mathcal{N}(0, \mathbf{I})$
for all t from T to 1 **do**
$\mu, \Sigma \leftarrow \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)$
$x_{t-1} \sim \mathcal{N}(\mu, (1+s)\Sigma)$
end for
return x_0

where $\overline{M} = 1 - M$, $\odot$ denotes the element-wise multiplication operation, and δ is the opacity in the image blending. To ensure that the generated anomalous regions are located in the foreground, we use an adaptive threshold-based binarization method for foreground segmentation, similar to methods used in [29, 35, 36]. Fig. 3a shows the images generated by SDAS under different anomaly strengths, while Fig. 3b compares the images with local anomaly regions synthesized by different methods. The larger the value of s, the greater the distribution difference between the generated image and the normal image, and the more obvious the abnormal region obtained after image blending. When s is very small, imperceptible abnormal regions can be synthesized. Compared with alternative synthesis methods, the anomalies generated by SDAS are more continuous and can have very realistic structural anomalies.

3.2. Anomaly-aware Features Selection

In this section, we introduce the Anomaly-aware Features Selection (AFS) module within RealNet, a self-supervised method for pre-trained feature selection, reducing feature dimensionality and eliminating pre-training bias, as well as managing reconstruction costs. Firstly, we define a set of N triplets $\{A_n, I_n, M_n\}_{n=1}^N$, where $A_n, I_n \in R^{h \times w \times 3}$ represent anomaly images synthesized by SDAS and original normal images, and $M_n \in R^{h \times w}$ represents the corresponding anomaly mask. We denote the pre-trained network as ϕ_k, and $\phi_k(A_n) \in R^{h_k \times w_k \times c_k}$ represents the kth layer pre-trained feature extracted from A_n, where c_k represents the number of channels. For the ith feature map, $\phi_{k,i}(A_n) \in R^{h_k \times w_k}$, AFS selects m_k feature maps for reconstruction ($m_k \leq c_k$). Specifically, the feature maps indexed by k are from ResNet-like architectures, such as ResNet50 [13] or WideResNet50 [41], where $k \in \{1, 2, 3, 4\}$ represent the last layer outputs of blocks with different spatial resolutions.

For the kth layer pre-trained features, we define the following AFS loss for evaluation of the ith feature map:

$$\mathcal{L}_{AFS}(\phi_{k,i}) = \frac{1}{N}\sum_{n=1}^{N}\|F([\phi_{k,i}(A_n)-\phi_{k,i}(I_n)]^2)-M_n\|_2^2 \tag{4}$$

where $F(\cdot)$ is a function that performs normalization operation and aligns the resolution of $[\phi_{k,i}(A_n)-\phi_{k,i}(I_n)]^2$ to M_n. Given the feature reconstruction process for anomalous images, we train a reconstruction network to infer $\phi_{k,i}(I_n)$ based on $\phi_{k,i}(A_n)$, which enables the detection and localization of anomalies through $[\phi_{k,i}(A_n)-\phi_{k,i}(I_n)]^2$. Ideally, $[\phi_{k,i}(A_n)-\phi_{k,i}(I_n)]^2$ should closely approximate M_n. The $\mathcal{L}_{AFS}(\phi_{k,i})$ represents the capability of $\phi_{k,i}$ in identifying anomalous regions. Due to the unavailability of real anomalous samples, we employ synthetic anomalies for feature selection. For the kth layer of pre-trained features, AFS selects m_k feature maps with the smallest $\mathcal{L}_{AFS}$ for reconstruction. We denote the AFS as $\varphi_k(\cdot)$, and $\varphi_k(A_n) \in R^{h_k\times w_k\times m_k}$, where $m_k \leq c_k$. We perform AFS on each layer of pre-trained features separately, and finally obtain selected multi-scale features $\{\varphi_1(A_n),...,\varphi_K(A_n)\}$. In this process, each layer's feature dimension $\{m_1,...,m_K\}$ serves as a set of hyperparameters. Specifically, for RealNet, AFS operation is performed only once on the pre-trained features of each layer, and the index of the selected feature maps is cached for subsequent training and inference.

AFS adaptively selects a subset of features from all available layers for anomaly detection, offering the following advantages compared to conventional methods [27, 32, 46] that select all features from partial layers: 1) AFS reduces feature redundancy within layers and mitigates pre-training bias, enhancing both feature representativeness and discriminability to improve anomaly detection performance. 2) AFS broadens the receptive field to enhance multi-scale anomaly detection capabilities. 3) AFS distinguishes the dimensions of pre-trained features from those employed for anomaly detection, ensuring efficient control over computational costs and flexible customization of the model size.

In RealNet, a set of reconstruction networks $\{G_1,...,G_K\}$ are designed to reconstruct the selected synthetic anomalous features $\{\varphi_1(A_n),...,\varphi_K(A_n)\}$ into the original image features $\{\varphi_1(I_n),...,\varphi_K(I_n)\}$ at various resolutions. The loss function $\mathcal{L}_{recon}$ is defined as:

$$\mathcal{L}_{recon}(A,I) = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\|G_k(\varphi_k(A_n))-\varphi_k(I_n)\|_2^2 \tag{5}$$

During the reconstruction process, we intentionally forgo aligning multi-scale features [27, 30, 38] to preserve optimal performance. This choice is motivated by the potential drawbacks associated with aligning low-resolution features through down-sampling, which could compromise the network's detection resolution and increase the risk of misidentifying anomalies. On the other hand, aligning high-resolution features using up-sampling may result in unnecessary feature redundancy, leading to elevated reconstruction costs. A detailed discussion on the reconstruction network architectures can be found in Appendix C.

3.3. Reconstruction Residuals Selection

In this section, we present the Reconstruction Residuals Selection (RRS) module. Reconstruction residuals are denoted as $\{E_1(A_n),...,E_K(A_n)\}$, where $E_k(A_n) = [\varphi_k(A_n) - G_k(\varphi_k(A_n))]^2$. To obtain the global reconstruction residual $E(A_n) \in R^{h'\times w'\times m'}$, we up-sample the low-resolution reconstruction residuals and concatenate them channel-wise, where $m' = \sum_{k=1}^{K} m_k$, $h' = max(h_1,...,h_K)$, and $w' = max(w_1,...,w_K)$.

The reconstruction residuals in $E(A_n)$ is obtained from the pre-trained features of reconstructing corresponding layer, and the features of the same resolution only have good ability to capture anomalies within a certain range. For instance, subtle low-level texture anomalies can be effectively captured exclusively by reconstruction residuals derived from low-level features. Therefore, RRS selects only a subset of reconstruction residuals that contain the most anomalous information for the anomaly score generation, to achieve the highest possible recall of anomalous regions.

Firstly, RRS performs GlobalMaxPooling (GMP) and GlobalAveragePooling (GAP) on $E(A_n)$ to obtain $E_{GMP}(A_n), E_{GAP}(A_n) \in R^{m'}$ respectively. The r largest elements in $E_{GMP}(A_n)$ and $E_{GAP}(A_n)$ are then used to index the positions of $E(A_n)$ and obtain $E_{max}(A_n,r), E_{avg}(A_n,r) \in R^{h'\times w'\times r}$, which respectively represent the Top$K$ reconstruction residuals with the highest maximum and average values. To avoid missed detections caused by inadequate resolution, reconstruction residuals with insufficient anomalous information are discarded in RRS.

As GMP and GAP respectively represent local and global properties spatially, E_{max} is more effective in capturing local anomalies in small areas, while E_{avg} focuses on selecting anomalies with large spans. Combining E_{max} and E_{avg} together can enhance the RRS's ability to capture anomalies of various scales. We define the RRS operator as $E_{RRS}(A_n,r) \in R^{h'\times w'\times r}$. $E_{RRS}(A_n,r)$ concatenates $E_{max}(A_n,r/2)$ and $E_{avg}(A_n,r/2)$. Finally, we feed the $E_{RRS}(A_n,r)$ into a discriminator, which maps the reconstruction residual to the image-level resolution, obtaining the final anomaly scores. The maximum value in anomaly scores is used as the image-level anomaly score. We use cross entropy loss $\mathcal{L}_{seg}(A,M)$ to supervise the training of

Table 1. Comparison of SIA with alternative anomaly synthesis approaches on the MVTec-AD dataset [3], employing Image AUROC (%), Pixel AUROC (%), and PRO (%) as evaluation metrics.

Category		SIA	DTD [5]	NSA [29]	CutPaste [18]
Texture	Carpet	(99.84, 99.19, 96.41)	(**100.0**, **99.27**, **96.96**)	(99.80, 98.60, 88.77)	(99.24, 98.42, 93.85)
	Grid	(**100.0**, 99.51, **97.28**)	(**100.0**, **99.57**, 97.14)	(**100.0**, 99.32, 91.31)	(**100.0**, 99.18, 92.53)
	Leather	(**100.0**, 99.76, 96.22)	(**100.0**, **99.77**, 96.41)	(**100.0**, 99.24, **96.85**)	(**100.0**, 99.41, 92.12)
	Tile	(99.96,**99.44**, **97.70**)	(**100.0**, 99.35, 95.27)	(**100.0**, 97.40, 86.45)	(99.86, 97.63, 84.39)
	Wood	(99.21, 98.22, 90.54)	(**99.65**, **98.28**, **91.23**)	(97.63, 93.30, 87.20)	(98.95, 95.29, 81.47)
	AVG	(99.80, 99.22, **95.63**)	(**99.93**, **99.25**,95.40)	(99.49, 97.57, 90.11)	(99.61, 97.99, 88.87)
Object	Bottle	(**100.0**, 99.30, **95.62**)	(**100.0**, 99.35, 95.57)	(**100.0**, **99.37**, 93.49)	(**100.0**, 99.14, 91.41)
	Cable	(99.19, **98.10**, **93.38**)	(98.95, 97.84, 90.36)	(**99.33**, 97.62, 93.26)	(96.35, 96.23, 86.05)
	Capsule	(**99.56**, **99.32**, 84.48)	(99.32, 99.19, 82.28)	(99.04, 99.27, **85.77**)	(98.48, 99.10, 79.55)
	hazelnut	(**100.0**, **99.68**, 93.14)	(**100.0**, 99.46, 93.46)	(**100.0**, 99.25, **94.41**)	(**100.0**, 99.03, 91.51)
	Metal Nut	(99.76, 98.58, 94.39)	(99.90, 98.58, **96.49**)	(**100.0**, **99.11**, 93.27)	(99.90, 98.03, 89.69)
	Pill	(**99.13**, **99.02**, 91.04)	(98.36, 98.88, 84.44)	(97.19, 98.28, **95.15**)	(97.22, 98.96, 86.48)
	Screw	(**98.83**, 99.45, 87.90)	(97.72, 99.36, 85.22)	(98.79, **99.62**, **93.74**)	(92.74, 98.53, 79.63)
	Toothbrush	(99.44, 98.71, **91.57**)	(99.44, 98.69, 90.87)	(**100.0**, **99.18**, 89.20)	(99.17, 98.85, 78.48)
	Transistor	(**100.0**, **98.00**, **92.92**)	(99.71, 97.15, 86.56)	(98.54, 95.67, 79.09)	(99.38, 96.32, 76.52)
	zipper	(99.82, **99.17**, **93.43**)	(99.68, 99.02, 88.77)	(**99.90**, 98.91, 93.05)	(99.61, 98.03, 92.26)
	AVG	(**99.57**, **98.93**, **91.79**)	(99.31, 98.75, 89.40)	(99.28, 98.63, 91.04)	(98.29, 98.22, 85.16)
AVG		(**99.65**, **99.03**, **93.07**)	(99.52, 98.92, 91.40)	(99.35, 98.28, 90.73)	(98.73, 98.14, 86.40)

Table 2. Comparison of RealNet with alternative anomaly detection methods on the MVTec-AD dataset [3].

Metric	*PatchCore* [27]	*SimpleNet* [19]	*FastFlow* [40]	DRAEM+SSPCAB [26]	DSR [43]	UniAD [38]	RD++ [32]	DeSTSeg [46]	DiffAD [45]	RealNet
Image AUROC	99.1	**99.6**	99.3	98.9	98.2	96.6	99.4	98.6	98.7	**99.6**
Pixel AUROC	98.1	98.1	98.1	97.2	-	96.6	98.3	97.9	98.3	**99.0**

discriminator. The overall loss of RealNet is:

$$\mathcal{L}(A, I, M) = \mathcal{L}_{recon}(A, I) + \mathcal{L}_{seg}(A, M) \tag{6}$$

3.4. Synthetic Industrial Anomaly Dataset

To facilitate the reuse of generated anomaly images by SDAS, we constructed the Synthetic Industrial Anomaly Dataset (SIA). SIA comprises anomaly images for 36 categories from four industrial anomaly detection datasets, including MVTec-AD [3], MPDD [16], BTAD [22], and VisA [47]. We generated 10,000 anomaly images with a resolution of 256×256 for each category, with anomaly strength s uniformly sampled between 0.1 and 0.2. SIA can be conveniently used for synthesizing anomaly images through image blending, as described in Eq. (3), and can serve as an effective alternative to the widely used DTD dataset [5].

4. Experiment

4.1. Experimental setup

Datasets. We conduct extensive evaluations on four datasets, including MVTec-AD [3], MPDD [16], BTAD [22], and VisA [47]. MVTec-AD [3] contains 5,354 images from 15 categories for industrial anomaly detection tasks, including 10 object categories and 5 texture categories. MPDD [16] contains 1,346 images from 6 types of industrial metal products with varying lighting conditions, non-uniform backgrounds, and multiple products in each image. Furthermore, the placement orientation, shooting distance, and position of the products are also varied. BTAD [22] contains images of 3 industrial products from the real world. VisA [47] is comprised of 9,621 normal images and 1,200 anomaly images from 12 categories. Certain categories demonstrate intricate structures, as exemplified by PCBs, while others consist of multiple objects that require detection, such as Capsules, thus rendering detection and localization a challenging task.

Metrics. To evaluate the performance of image-level anomaly detection, we use the Area Under the Receiver Operator Curve (AUROC) metric, as in previous works [3, 16, 22, 47]. For pixel-level anomaly location, we use Pixel AUROC and Per Region Overlap (PRO) [4].

Implementation details. We evaluate RealNet on four datasets with consistent network architectures and hyperparameters, without specific tuning for individual categories. We use a WideResNet50 [41] pre-trained on ImageNet [9] as the backbone. In AFS, we set the dimension of pre-trained feature of each layer to $\{256, 512, 512, 256\}$ for reconstruction. For RRS, $1/3$ of the reconstruction residuals are reserved to generate the final anomaly scores. For SDAS, we train the diffusion model following [10] and use the SIA dataset for anomaly synthesis. Both SDAS and anomaly detection are performed at a resolution of 256×256 without center cropping, with a batch size of 16, and we use 64 batches of synthetic anomaly images for AFS. More details can be found in Appendix B.

Table 3. Comparison of SIA with DTD [5] and CutPaste [18] on the MPDD dataset [16], employing Image AUROC (%), Pixel AUROC (%), and PRO (%) as evaluation metrics.

Category	SIA	DTD [5]	CutPaste [18]
Bracket Black	(**94.95**, **99.27**, 87.10)	(89.49, 98.90, **88.57**)	(66.42, 96.67, 56.53)
Bracket Brown	(**96.83**, **97.81**, **94.36**)	(92.99, 97.35, 92.64)	(95.48, 97.54, 55.17)
Bracket White	(**88.78**, 97.44, **84.00**)	(86.67, **98.59**, 77.08)	(88.44, 96.51, 64.32)
Connector	(**100.0**, 97.46, **84.79**)	(99.05, 97.76, 65.91)	(99.05, **98.47**, 74.05)
Metal Plate	(**100.0**, 99.28, **94.44**)	(**100.0**, **99.35**, 93.78)	(99.95, 98.83, 92.69)
Tubes	(**97.51**, 97.94, 93.29)	(92.62, **99.01**, **96.49**)	(91.49, 98.09, 92.99)
AVG	(**96.35**, 98.20, **89.66**)	(93.47, **98.49**, 85.75)	(90.14, 97.69, 72.63)

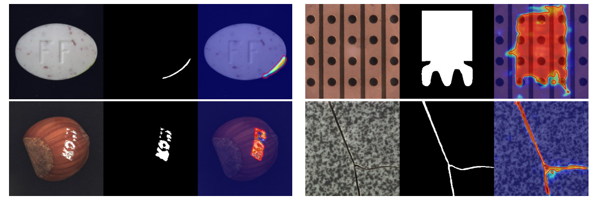

Figure 4. Qualitative results of RealNet on the MVTec-AD dataset [3]. Within each group, from left to right, are the anomaly image, ground-truth, and predicted anomaly score.

Table 4. Comparison of RealNet with alternative anomaly detection methods on the MPDD dataset [16].

Metric	*PatchCore* [27]	*CFlow* [12]	*PaDiM* [7]	*SPADE* [6]	DAGAN [31]	Skip-GANomaly [1]	RealNet
Image AUROC	82.1	86.1	74.8	77.1	72.5	64.8	**96.3**
Pixel AUROC	95.7	97.7	96.7	95.9	83.3	82.2	**98.2**

4.2. Anomaly detection on MVTec-AD

We train RealNet using SIA and alternative anomaly synthetic methods on the MVTec-AD dataset [3], to evaluate the model's performance in anomaly detection and localization. These methods include: 1) **DTD** [5]: This method utilizes the DTD dataset [5] to blend images with generated anomalous textures, and the data augmentation strategy in [42] is employed during the blending process. 2) **NSA** [29]: This method employs Poisson image editing [23] to seamless image editing, following parameter setting in [29]. 3) **CutPaste** [18]: This method involves random image cropping and pasting to synthesize anomaly regions.

The experimental results are shown in Tab. 1. SDAS demonstrates flexibility in controlling the anomaly strength and generates synthetic anomalies with multiple anomaly patterns, especially for the object categories, where it achieves the best detection and localization performance. Compared to other methods, SDAS is not constrained by data augmentation rules or external data, enabling the synthesis of more natural and rich functional anomalies, as shown in Fig. 3. RealNet trained using SIA achieves remarkable performance on the MVTec-AD dataset [3], with an Image AUROC of 99.65%, a Pixel AUROC of 99.03%, and a PRO score of 93.07%. Fig. 4 presents the qualitative anomaly localization results of RealNet on the MVTec-AD dataset [3]. The method exhibits remarkable pixel-level anomaly localization, proficiently identifying diverse anomaly patterns at various scales. Furthermore, RealNet can achieve a rapid inference speed of 31.93 FPS when using a single Nvidia GeForce RTX 3090, and it can perform inference using only 4GB of GPU memory. A detailed computational efficiency analysis can be found in Appendix C.

We also compare RealNet with several state-of-the-art anomaly detection methods, and the results are shown in Tab. 2. Built on the same pre-trained network, RealNet outperforms the state-of-the-art alternatives, including Deep Feature Embedding-Based method (PatchCore [27] and SimpleNet [19]) and the NF-Based method (FastFlow [40]). When compared to previous reconstruction-based methods, RealNet achieves significant performance improvement.

4.3. Anomaly detection on MPDD

We evaluate RealNet on MPDD dataset [16] with SIA, DTD [5], and CutPaste [18], and the results are shown in Tab. 3. Notably, RealNet trained with SIA achieves a significant improvement of 2.88% in Image AUROC over DTD [5]. Tab. 4 shows the Image AUROC and the Pixel AUROC of RealNet and other methods on the MPDD dataset [16]. RealNet achieves an Image AUROC of 96.3%, surpassing the current best performance (CFlow [12]) by 10.2%, even without any dataset-specific tuning.

4.4. Anomaly detection on other benchmarks

To comprehensively assess the effectiveness of RealNet, we conduct experiments on the BTAD [22] and VisA [47] datasets. On the VisA dataset [47], characterized by complex structures and multiple detection objects, RealNet employing SIA yields a significant performance improvement, achieving an Image AUROC of 97.8% and a Pixel AUROC of 98.8%. In the case of the BTAD dataset [22], RealNet with SIA achieves competitive results, securing an Image AUROC of 96.1% and a Pixel AUROC of 97.9%. Detailed results can be found in Appendix C.

4.5. Ablation studies

To evaluate the effectiveness of each module of RealNet, we conduct comprehensive ablation studies on MVTec-AD dataset [3]. First, we evaluate the impact of AFS and RRS on the performance of RealNet.

W/O AFS: We replace AFS with two alternative dimensionality reduction methods, namely Random Dimensionality Reduction (RDR) [7] and Random Linear Projections

Table 5. Ablation studies of RealNet on the MVTec-AD dataset [3].

(a) The impact of AFS and RRS on RealNet.

	AFS	RRS	Image AUROC	Pixel AUROC	PRO
1	-	-	94.46 / 95.67	93.38 / 95.84	79.81 / 82.26
2	✓	-	96.86	96.32	84.13
3	-	✓	99.39 / 99.09	98.66 / 98.22	92.01 / 88.38
4	✓	✓	**99.65**	**99.03**	**93.07**

(b) The impact of anomaly strengths on RealNet.

Metric	$s = 0$	$s = 0.1$	$s = 0.2$	$s \in [0.1, 0.2]$
Image AUROC	99.35	**99.65**	99.61	**99.65**
Pixel AUROC	98.85	98.96	98.95	**99.03**
PRO	91.80	92.16	89.36	**93.07**

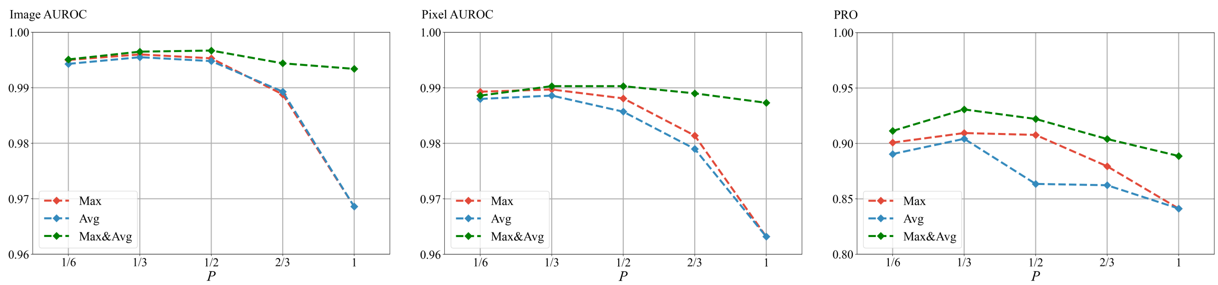

Figure 5. Performance of RealNet on MVTec-AD dataset [3] under various modes of reconstruction residuals selection (`Max`, `Avg`, and `Max&Avg`) and varying retention ratio P of reconstruction residuals.

Reduction (RLPR) [27, 34]. RDR randomly selects some dimensional features from high-dimensional features, while RLPR employs an untrained linear transformation layer for linear projection. We report the results of our RealNet with RDR and RLPR, respectively, as shown in the experiments 1 and 3 in Tab. 5a. **W/O RRS:** We feed the global reconstruction residual $E(A_n)$ into the discriminator to generate anomaly scores, and the results are shown in the experiments 1 and 2 in Tab. 5a.

As indicated by the ablation results in Tab. 5a, RRS contributes significantly to performance improvement. Using all the reconstruction residuals to generate anomaly scores, reconstruction residuals lacking of anomaly information can lead to missed anomaly regions, resulting in a significant decrease in anomaly detection performance. Furthermore, AFS yields better anomaly detection results compared to RDR and RLPR. A straightforward visualization result about AFS is provided in Appendix D.

We further investigate the impact of anomaly strength s in SDAS, and the results are shown in Tab. 5b. When s equals 0, SDAS generates normal images in high probability density regions. Blending images may introduce false positive anomaly regions, which lowers the reconstruction difficulty and confuses the discriminator, leading to suboptimal performance. Conversely, when s is too large, the synthetic anomaly images deviate from the true distribution of anomalous images, causing RealNet's performance to deteriorate. Our findings indicate that uniformly sampling s within a specific range serves as a robust approach for generating anomalous images. This method enables RealNet to cover a wider range of anomalous patterns, ultimately improving the overall anomaly detection performance.

In Fig. 5, we report the impact of different RRS modes and retention ratios on the performance of RealNet. Since the setting of r is related to m_k, we introduce the retention ratio P, defined as: $P = \frac{r}{\sum_{k=1}^{K} m_k}$. Compared to `Max` and `Avg` modes, `Max&Avg` mode demonstrates superior robustness in detecting anomalies across various scales. At equal retention rates, the `Max&Avg` mode discards more reconstruction residuals lacking anomalous information than the `Max` and `Avg` modes, mitigating performance degradation and further emphasizing the effectiveness of the `Max&Avg` mode in enhancing RealNet's anomaly detection capabilities. More ablation experiments and analysis can be found in Appendix C.

5. Conclusion

In this work, we introduce RealNet, an innovative self-supervised anomaly detection framework. Our approach integrates three core components: Strength-controllable Diffusion Anomaly Synthesis (SDAS), Anomaly-aware Features Selection (AFS), and Reconstruction Residuals Selection (RRS). These components synergistically contribute to RealNet, enabling effective exploitation of large-scale pre-trained models in anomaly detection while keeping the computational overhead within a reasonably low and acceptable range. RealNet provides a flexible foundation for future research in anomaly detection utilizing pre-trained feature reconstruction techniques. Through extensive experiments, we illustrate RealNet's proficiency in addressing diverse real-world anomaly detection challenges.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant 62177034 and Grant 61972046.

References

[1] Samet Akçay, Amir Atapour-Abarghouei, and Toby P Breckon. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 2, 7

[2] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pages 161–169. Springer, 2019. 2

[3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec-ad: A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1, 2, 6, 7, 8

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4183–4192, 2020. 6

[5] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3606–3613, 2014. 2, 6, 7

[6] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 7

[7] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV*, pages 475–489. Springer, 2021. 2, 7

[8] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9737–9746, 2022. 1, 2

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2, 6

[10] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 6

[11] 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*, pages 571–578, 2023. 3

[12] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 98–107, 2022. 7

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[14] Lars Heckler, Rebecca König, and Paul Bergmann. Exploring the importance of pretrained feature extractors for unsupervised anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2916–2925, 2023. 2

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1, 3

[16] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 66–71. IEEE, 2021. 2, 6, 7

[17] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 3

[18] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021. 1, 2, 6, 7

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

[20] Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus Robert Muller. Explainable deep one-class classification. In *International Conference on Learning Representations*, 2021. 2

[21] Fanbin Lu, Xufeng Yao, Chi-Wing Fu, and Jiaya Jia. Removing anomalies as noises for industrial defect localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16166–16175, 2023. 2

[22] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, pages 01–06. IEEE, 2021. 2, 6, 7

[23] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, pages 313–318. 2003. 2, 7

[24] Ken Perlin. An image synthesizer. *ACM Siggraph Computer Graphics*, 19(3):287–296, 1985. 4

[25] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *Image Analysis and Processing–ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part II*, pages 394–406. Springer, 2022. 2

[26] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13576–13586, 2022. 6

[27] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 2, 5, 6, 7, 8

[28] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings*, pages 146–157. Springer, 2017. 2

[29] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 474–489. Springer, 2022. 1, 2, 4, 6, 7

[30] Yong Shi, Jie Yang, and Zhiquan Qi. Unsupervised anomaly segmentation via deep feature reconstruction. *Neurocomputing*, 424:9–22, 2021. 1, 2, 5

[31] Ta-Wei Tang, Wei-Han Kuo, Jauh-Hsiang Lan, Chien-Fang Ding, Hakiem Hsu, and Hong-Tsu Young. Anomaly detection neural network with dual auto-encoders gan and its industrial inspection applications. *Sensors*, 20(12):3336, 2020. 7

[32] 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*, pages 24511–24520, 2023. 2, 5, 6

[33] Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 650–656, 2022. 2

[34] Jiang Xi, Jianlin Liu, Jinbao Wang, Qiang Nie, WU Kai, Yong Liu, Chengjie Wang, and Feng Zheng. Softpatch: Unsupervised anomaly detection with noisy data. In *Advances in Neural Information Processing Systems*. 8

[35] 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. 4

[36] Xincheng Yao, Ruoqi Li, Jing Zhang, Jun Sun, and Chongyang Zhang. 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*, pages 24490–24499, 2023. 4

[37] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2

[38] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. In *Advances in Neural Information Processing Systems*, 2022. 1, 2, 5, 6

[39] Sanyapong Youkachen, Miti Ruchanurucks, Teera Phatrapomnant, and Hirohiko Kaneko. Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing. In *2019 10th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, pages 1–5. IEEE, 2019. 2

[40] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 2, 6, 7

[41] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press, 2016. 4, 6

[42] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem: A discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 1, 2, 4, 7

[43] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr: A dual subspace re-projection network for surface anomaly detection. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 539–554. Springer, 2022. 1, 2, 6

[44] Hui Zhang, Zuxuan Wu, Zheng Wang, Zhineng Chen, and Yu-Gang Jiang. Prototypical residual networks for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16281–16291, 2023. 1

[45] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6782–6791, 2023. 2, 6

[46] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3914–3923, 2023. 1, 2, 5, 6

[47] 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: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXX*, pages 392–408. Springer, 2022. 2, 6, 7