Dissolving Is Amplifying: Towards Fine-Grained Anomaly Detection

Jian Shi[1], Pengyi Zhang[2], Ni Zhang[2], Hakim Ghazzai[1], and Peter Wonka[1]

[1] King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
{jian.shi, hakim.ghazzai, peter.wonka}@kaust.edu.sa
[2] NEC Laboratories China, Beijing, China
{zhang_pengyi,zhangni_nlc}@nec.cn

Abstract. Medical imaging often contains critical fine-grained features, such as tumors or hemorrhages, which are crucial for diagnosis yet potentially too subtle for detection with conventional methods. In this paper, we introduce *DIA*, dissolving is amplifying. DIA is a fine-grained anomaly detection framework for medical images. First, we introduce *dissolving transformations*. We employ diffusion with a generative diffusion model as a dedicated feature-aware denoiser. Applying diffusion to medical images in a certain manner can remove or diminish fine-grained discriminative features. Second, we introduce an *amplifying framework* based on contrastive learning to learn a semantically meaningful representation of medical images in a self-supervised manner, with a focus on fine-grained features. The amplifying framework contrasts additional pairs of images with and without dissolving transformations applied and thereby emphasizes the dissolved fine-grained features. DIA significantly improves the medical anomaly detection performance with around 18.40% AUC boost against the baseline method and achieves an overall SOTA against other benchmark methods. Our code is available at `https://github.com/shijianjian/DIA.git`.

1 Introduction

Anomaly detection aims to detect exceptional data instances that significantly deviate from normal data. A popular application is the detection of anomalies in medical images, where these anomalies often indicate a form of disease or medical problem. In the medical field, anomalous data is scarce and diverse, so anomaly detection is commonly modeled as semi-supervised anomaly detection. This means that anomalous data is not available during training, and the training data contains only the "normal" class.[1] Traditional anomaly detection methods include one-class methods (*e.g.* One-class SVM [14]), reconstruction-based methods (*e.g.* AutoEncoders [57]), and statistical models (*e.g.* HBOS [22]). However, most anomaly detection methods suffer from a low recall rate, meaning that many normal samples are wrongly reported as anomalies while true yet

---

[1] Some early studies refer to training with only normal data as unsupervised anomaly detection. However, we follow [36, 37] and other newer methods and use the term semi-supervised.

sophisticated anomalies are missed [37]. Notably, due to the nature of anomalies, the collection of anomaly data can hardly cover all anomaly types, even for supervised classification-based methods [38]. An inherited challenge is the inconsistent behavior of anomalies, which varies without a concrete definition [9, 55]. Thus, identifying unseen anomalous features without requiring prior knowledge of anomalous feature patterns is crucial to anomaly detection applications.

In order to identify unseen anomalous features, many studies leveraged data augmentations [21, 60] and adversarial features [2] to emphasize various feature patterns that deviate from normal data. This field attracted more attention after incorporating Generative Adversarial Networks (GANs) [23], including [1, 2, 45, 46, 51, 52, 65], to enlarge the feature distances between normal and anomalous features through adversarial data generation methods. Furthermore, some studies [35, 40, 49] explored the use of GANs to deconstruct images to generate out-of-distribution data for obtaining more varied anomalous features. Inspired by the recent successes of contrastive learning [8, 10–13, 24, 28], contrastive-based anomaly detection methods such as Contrasting Shifted Instances (CSI) [53] and mean-shifted contrastive loss [42] improve upon GAN-based methods by a large margin. The contrastive-based methods fit the anomaly detection context well, as they are able to learn robust feature encoding without supervision. By comparing the feature differences between positive pairs (*e.g.* the same image with different views) and negative pairs (*e.g.* different images w/wo different views) without knowing the anomalous patterns, contrastive-based methods achieved outstanding performance in many general anomaly detection tasks [42, 53]. However, given the low performance in experiments in Sec. 4, those methods are less effective for medical anomaly detection. We suspect that contrastive learning in conjunction with traditional data augmentation methods (*e.g.* crop, rotation) cannot focus on fine-grained features and only identifies coarse-grained feature differences well (*e.g.* car vs. plane). As a result, medical anomaly detection remains challenging because models struggle to recognize these fine-grained, inconspicuous, yet important anomalous features that manifest differently across individual cases. These features are critical for identifying anomalies but can be subtle and easily overlooked. Thus, in this work, we investigate the principled question: *how to emphasize the fine-grained features for fine-grained anomaly detection?*

Our method. This paper dissects the complex feature patterns within medical datasets into two distinct categories: discriminative and non-discriminative features. Discriminative features are commonly unique and fine-grained characteristics that allow for the differentiation of individual data samples, serving as critical markers for identification and classification. Conversely, non-discriminative features encompass the shared patterns that define the general semantic context of the dataset, offering a backdrop against which the discriminative features stand out. To aid the learning of fine-grained discriminative feature patterns, we propose an intuitive contrastive learning strategy to compare an image against its transformed version with fewer discriminative features to emphasize the removed fine-grained details. We introduce *dissolving transforma-*

(a) Input Images   **(b)** $t = 50$ **(c)** $t = 100$ **(d)** $t = 200$ **(e)** $t = 400$
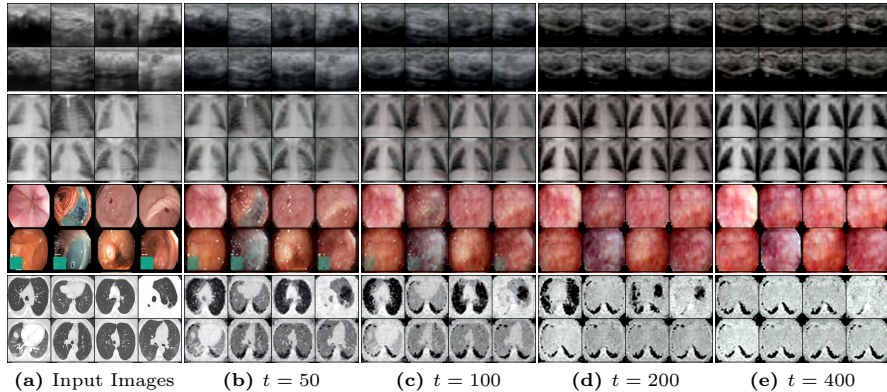
Fig. 1: Dissolving Transformations. Figs. 1b to 1e show how the fine-grained features are dissolved (removed or suppressed). This effect is stronger as the time step t is increased from left to right. In the extreme case, in Fig. 1e, different input images become very similar or almost identical depending on the dataset. We show results for four datasets from top to bottom.

tions based on pre-trained diffusion models, that leverage the individual reverse diffusion steps within the diffusion models to function as feature-aware denoisers, to remove or suppress fine-grained discriminative features from an input image. We also introduce the framework *DIA*, dissolving is amplifying, that leverages the proposed *dissolving transformations*. DIA is a contrasting learning framework. Its enhanced understanding of fine-grained discriminative features stems from a loss function that contrasts images that have been transformed with dissolving transformations to images that have not. On six medical datasets, our method obtained roughly an 18.40% AUC boost against the baseline method and achieved the overall SOTA compared to existing methods for fine-grained medical anomaly detection. Key contributions of *DIA* include:

– **Conceptual Contribution.** We propose a novel strategy that enhances the detection of fine-grained, subtle anomalies without requiring pre-defined anomalous feature patterns, by emphasizing the differences between images and their feature-dissolved counterparts.
– **Technical Contribution 1.** We introduce *dissolving transformations* to dissolve the fine-grained features of images. It performs semantic feature dissolving via the reverse process of diffusion models as described in Fig. 1.
– **Technical Contribution 2.** We present an *amplifying strategy* for self-supervised fine-grained feature learning, leveraging a *fine-grained NT-Xent loss* to learn fine-grained discriminative features.

2   Related Work

2.1   Synthesis-based Anomaly Detection

As [37, 41] indicated, semi-supervised anomaly detection methods dominated this research field. These methods utilized only normal data whilst training.

With the introduction of GANs [23], many attempts have been made to bring GANs into anomaly detection. Here, we roughly categorize current methods to *reconstructive synthesis* that increases the variation of normal data, and *deconstructive synthesis* that generates more anomalous data.

Reconstructive Synthesis. Many studies [6, 64] focused on synthesizing various in-distribution data (*i.e.* normal data) with synthetic methods. For anomaly detection tasks, earlier works such as AnoGAN [50] learn normal data distributions with GANs that attempt to reconstruct the most similar images by optimizing a latent noise vector iteratively. With the success of Adversarial Auto Encoders (AAE) [33], some more recent studies combined AutoEncoders and GANs together to detect anomalies. GANomaly [1] further regularized the latent spaces between inputs and reconstructed images, and then some following works improved it with more advanced generators such as UNet [2] and UNet++ [15]. AnoDDPM [58] replaced GANs with diffusion model generators and stated the effectiveness of noise types for medical images (i.e., Simplex noise is better than Gaussian noise). In general, most of the *reconstructive synthesis* methods aim to improve normality feature learning despite the awareness of abnormalities, which impedes the model from understanding the anomaly feature patterns.

Deconstructive Synthesis. Due to the difficulties of data acquisition and to protect patient privacy, getting high-quality, balanced datasets in the medical field is difficult [29]. Thus, *deconstructive synthesis* methods are widely applied in medical image domains, such as X-ray [48], lesion [20], and MRI [27]. Recent studies tried to integrate such negative data generation methods into anomaly detection. G2D [40] proposed a two-phased training to train an anomaly image generator and then an anomaly detector. Similarly, ALGAN [35] proposed an end-to-end method that generates *pseudo-anomalies* during the training of anomaly detectors. Such GAN-based methods deconstruct images to generate *pseudo-anomalies*, resulting in unrealistic anomaly patterns, though multiple regularizers are applied to preserve image semantics. Unlike most works to synthesize novel samples from noises, we dissolve the fine-grained features on input data. Our method, therefore, learns the fine-grained instance feature patterns by comparing samples against their feature-dissolved counterparts. Benefiting from the step-by-step diffusing process of diffusion models, our proposed *dissolving transformations* can provide fine control over feature dissolving levels.

2.2 Contrastive-based Anomaly Detection

To improve anomaly detection performances, previous studies such as [19, 56] explored the discriminative feature learning to reduce the needs of labeled samples for supervised anomaly detection. More recently, GeoTrans [21] leveraged geometric transformations to learn discriminative features, which significantly improved anomaly detection abilities. ARNet [60] attempted to use embedding-guided feature restoration to learn more semantic-preserving anomaly features. Specifically, contrastive learning methods [8, 10–13, 24, 28] are proven to be promising in unsupervised representation learning. Inspired by the recent integration [16, 42, 53] of contrastive learning and anomaly detection tasks, we
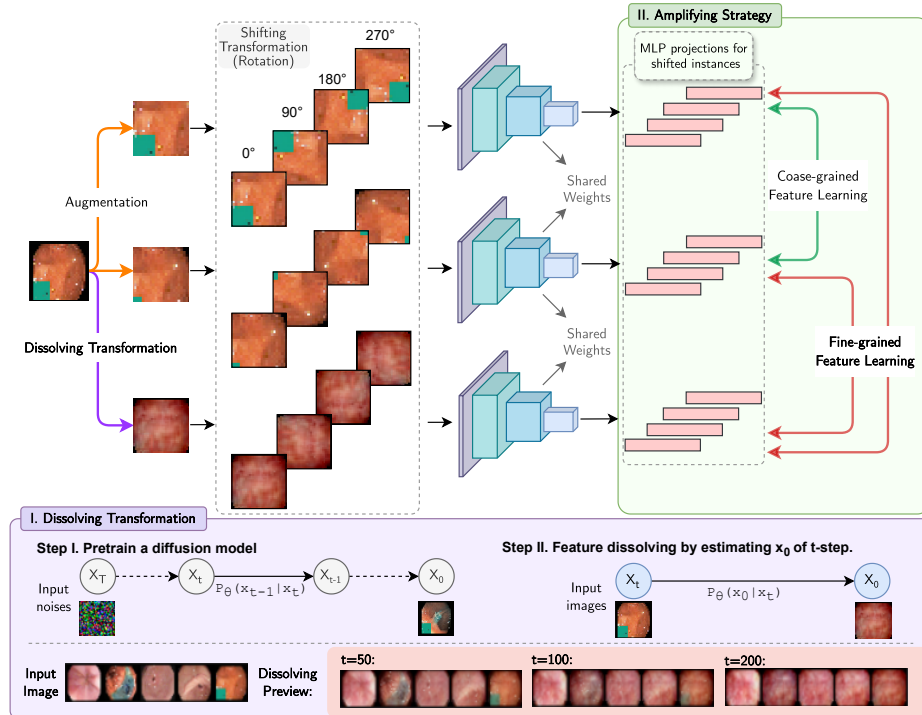
Fig. 2: An overview of the DIA framework as applied to the Kvasir-polyp dataset. (I) With a pretrained diffusion model, we perform feature-aware dissolving transformations on an image x. This process estimates the denoised version x_0 of x at a given time step t, resulting in a feature-dissolved image $\hat{x}$. As t increases, $\hat{x}$ progressively loses its fine-grained discriminative features, highlighting the dissolving effect of removing discriminative image features. (II) Given images, we generate transformed versions with augmentations and dissolving transformation. We form positive and negative pairs as described in Sec. 3.2. Our framework particularly learns fine-grained features by contrasting between original images and their feature-dissolved counterparts.

propose to construct negative pairs of a given sample and its corresponding *feature-dissolved* samples in a contrastive manner to enhance the awareness of fine-grained discriminative features for medical anomaly detection.

3   Methodology

This section introduces DIA (Dissolving Is Amplifying), a method curated for fine-grained anomaly detection for medical imaging. DIA is a self-supervised method based on contrastive learning, as illustrated in Fig. 2. DIA learns representations that can distinguish fine-grained discriminative features in medical images. First, DIA employs a dissolving strategy based on *dissolving transformations* (Sec. 3.1). The dissolving transformations can remove or deemphasize fine-grained discriminative features. Second, DIA uses the amplifying framework

described in Sec. 3.2 to contrast images that have been transformed with and without dissolving transformations. We use the term amplifying framework as it amplifies the representation of fine-grained discriminative features.

3.1   Dissolving Strategy

We introduce *dissolving transformations* to create negative examples in a contrastive learning framework. The dissolving transformations are achieved by pretrained diffusion models. The output image maintains a similar structure and appearance to the input image, but several fine-grained discriminative features unique to the input image are removed or attenuated. Unlike the regular diffusion process that starts with pure noise, we initialize with the input image without adding noise. As depicted in Fig. 1, *dissolving transformations* progressively remove fine details from various datasets as diffusion time steps t increase.

To recap, diffusion models consist of forward and reverse processes, each performed over T time steps. The forward process q gradually adds noise to an image x_0 for T steps to obtain a pure noise image x_T, whereas the reverse process p aims at restoring the starting image x_0 from x_T. In particular, we sample an image $x_0 \sim q(x_0)$ from a real data distribution $q(x_0)$, then add noise at each step t with the forward process $q(x_t|x_{t-1})$, which can be expressed as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} \cdot x_{t-1}, \beta_t \cdot \mathrm{I}), \tag{1}$$

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}), \tag{2}$$

where β_t represents a known variance schedule that follows $0 < \beta_1 < \beta_2 < \cdots < \beta_T < 1$. Afterwards, the reverse process removes noise starting at $p(x_T) = \mathcal{N}(x_T; 0, \mathrm{I})$ for T steps. Let θ be the network parameters:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)), \tag{3}$$

where μ_θ and Σ_θ are the mean and variance conditioned on step number t.

The proposed *dissolving transformations* are based on Eq. (3). Instead of generating images by progressive denoising, we apply reverse diffusion in a single step directly on an input image. Essentially, we set $x_t = x$ in Eq. (3), where x is the input image. We then compute an approximated state x_0 and denote it as $\hat{x}_{t \to 0}$ to make it clear that the equation below is parameterized by the time step t. By reparametrizing Eq. (3), $\hat{x}_{t \to 0}$ can be obtained by:

$$\hat{x}_{t \to 0} = \sqrt{\frac{1}{\bar{\alpha}_t}} \cdot x - \sqrt{\frac{1}{\bar{\alpha}_t} - 1} \cdot \epsilon_\theta(x, t), \quad \bar{\alpha}_t := \Pi_{s=1}^t \alpha_s \text{ and } \alpha_t := 1 - \beta_t, \tag{4}$$

where ϵ_θ is a function approximator (*e.g.* UNet) to predict the corresponding noise from x. Since a greater value of t leads to a higher variance β_t, $\hat{x}_{t \to 0}$ is expected to remove more of the "noise" if t is large. In our context, we do not remove "noise" but discriminative features. If t is small, the removed discriminative features are more fine-grained. If t is larger, larger discriminative features may be removed. See Fig. 1 and Sec. 6 for examples and an in-depth discussion.

3.2   Amplifying Framework

We propose a novel contrastive learning framework to enhance the awareness of the fine-grained image features by integrating the proposed *dissolving transformations*. In anomaly detection, the efficacy of contrastive learned features can be enhanced by applying *shifting transformations* [53]. A typical example is using significant rotations, which alters the distribution of the data based on the orientation of the transformed images. For instance, images rotated by 90 degrees are assimilated into the same distribution, whereas images subjected to a 180-degree rotation diverge from this distribution. However, this improved contrastive feature learning technique does not come with a fine-grained feature learning mechanism, resulting in low performances on fine-grained anomaly detection tasks. We introduce feature-dissolved samples to augment the process of fine-grained feature learning. The feature-dissolved samples present significant differences from the original data, despite both sets belonging to the same shifting distributions. In particular, we aim to enforce the model to focus on fine-grained features by emphasizing the differences between images with and without *dissolving transformations*.

In our amplifying framework, we employ three types of transformations: *shifting transformations* (*e.g.* large rotations), *non-shifting transformations* (*e.g.* color jitter, random resized crop, and grayscale), and *dissolving transformations*. Our contrastive learning framework uniquely applies these transformations to input images through $3K$ distinct processes. The first $2K$ transformation branches are dedicated to coarse-grained feature learning, focusing on broader, more general features of the data. Conversely, the final K transformations are specifically tailored for fine-grained feature learning. This is accomplished by contrasting the transformed images against non-dissolved data samples, thereby enhancing the model's ability to discern subtle differences within the data. This approach not only broadens the scope of feature extraction but also significantly improves the model's precision in identifying nuanced patterns and anomalies.

Transformation Branches We use a set $\mathcal{S}$ of K different *shifting transformations*. This set contains only fixed (non-random) transformations and starts from the identity I so that $\mathcal{S} := \{S_0 = I, S_1, \ldots, S_{K-1}\}$. With input image x, we obtain $S_1(x), \ldots, S_{K-1}(x)$ as shifted images that strongly differ from the in-distribution samples $S_0(x) = x$. Each of these K shifted images then passes through multiple non-shifting transformations $\in \mathcal{T}$. This yields the set of combined transformations $O := \{O_0, O_1, \ldots, O_{K-1}\}$ and $O_k = \mathcal{T} \circ S_k$. With a slight abuse of notations, we use $\mathcal{T}$ as a sequence of random non-shifting transformations. This process is then repeated a second time, yielding another transformation set $\mathcal{O}'$. We also refer to $\mathcal{O}$ and $\mathcal{O}'$ as two augmentation branches. Each image is therefore transformed $2K$ times, K times in each augmentation branch. All transformations have supposedly different randomly sampled non-shifting transformations, but $O_i(x)$ and $O'_j(x)$ share the same *shifting transformation* if $i = j$. The introduced *dissolving transformations* serves as the third augmentation branch, denoted as $\mathcal{A} := \{A_0, \ldots, A_{K-1}\}$. The *dissolving transformations*

branch outputs transformations of the form:

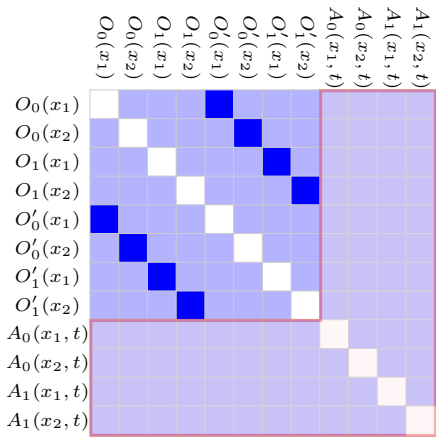$$A_k = \mathcal{T} \circ S_k \circ \mathcal{D} \tag{5}$$

where $\mathcal{T}$ is a sequence of random non-shifting transformations, S_k is a *shifting transformation*, and $\mathcal{D}$ is a randomly sampled *dissolving transformation*. In summary, this yields $3K$ transformations of each image, K in each of the three augmentation branches.

Fine-grained Contrastive Learning The goal of contrastive learning is to transform input images into a semantically meaningful feature representation. It is achieved by bringing similar examples (*i.e.* positive pairs) closer and pushing dissimilar examples (*i.e.* negative pairs) apart. To emphasize fine-grained features, an inherent strategy is to create negative pairs, where an image is contrasted with its transformed version with less fine-grained details, thereby enhancing the model's focus on these subtle distinctions.

For a single image, we have $3K$ different transformations. With B different images in a batch, yielding $3K \cdot B$ images that are considered jointly. For all possible pairs of images, they can either be a *negative pair*, a *positive pair*, or not be considered in the loss function. We relegate the explanation to an illustration in Fig. 3. In the top left quadrant of the matrix, we can see the design choices of what constitutes a positive and a negative pair inherited from [53], based on the *NT-Xent* loss [10]. The region highlighted in red, is our proposed design for the new *negative pairs* for *dissolving transformations*. The purpose of these newly introduced *negative pairs* is to learn a representation that can better distinguish between fine-grained semantically meaningful features. The contrastive loss for each image sample can be computed as follows:

Fig. 3: Visualization of the target similarity matrix ($K = 2$ with two samples in a batch). The white, blue, and lavender blocks denote the excluded, positive, and negative pairs, respectively. The red area contains the newly introduced negative pairs with dissolving transformations.

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{3N} \mathbb{1}_{k,i} \cdot (\exp(\text{sim}(z_i, z_k))/\tau)}, \quad \mathbb{1}_{k,i} = \begin{cases} 0 & i = k, \\ 1 & otherwise, \end{cases} \tag{6}$$

where N is the number of samples (*i.e.* $N = B \cdot K$), $sim(z, \hat{z}) = z \cdot \hat{z}/||z||||\hat{z}||$, and τ is a temperature hyperparameter to control the penalties of negative samples.

As mentioned, the positive pairs are selected from $O_i(\cdot)$ and $O'_j(\cdot)$ branches only when $i = j$. The proposed *feature-amplified NT-Xent* loss can therefore be expressed as:

$$\mathcal{L}_{con} = \frac{1}{3BK} \frac{1}{|\{x_+\}|} \sum \ell_{i,j} \cdot \begin{cases} 0 & \mathbb{1}_{i,j} \in \{x_-\} \\ 1 & \mathbb{1}_{i,j} \in \{x_+\} \end{cases}, \tag{7}$$

where $\{x_+\}$ and $\{x_-\}$ denote the positive and negative pairs, and $|\{x_+\}|$ is the number of positive pairs.

Additionally, an auxiliary softmax classifier f_θ is used to predict which shifting transformation is applied for a given input x, resulting in $p_{cls}(y^S|x)$. With the union of non-dissolving and dissolving transformed samples $\mathcal{X}_{\mathcal{S} \cup \mathcal{A}}$, the classification loss is defined as:

$$\mathcal{L}_{cls} = \frac{1}{3B} \frac{1}{K} \sum_{k=0}^{K-1} \sum_{\hat{x} \in \mathcal{X}_{\mathcal{S} \cup \mathcal{A}}} -\log p_{cls}(y^S|\hat{x}). \tag{8}$$

The final training loss is hereby defined as:

$$\mathcal{L}_{DIA} = \mathcal{L}_{con} + \gamma \cdot \mathcal{L}_{cls}, \tag{9}$$

where γ is set to 1 in this work.

3.3　The Score functions

During inference, we adopt an anomaly score function that consists of two parts: (1) s_{con} sums the anomaly scores over all shifted transformations, in addition to (2) s_{cls} sums the confidence of the shifting transformation classifier. For the kth shifting transformation, given an input image x, training example set $\{x_m\}$, and a feature extractor c, we have:

$$s_{con}(\tilde{x}, \{\tilde{x}_m\}) = \max_m \; \mathrm{sim}(c(\tilde{x}_m), c(\tilde{x})) \cdot ||c(\tilde{x})||, s_{cls}(\tilde{x}) = W_k f_\theta(\tilde{x}), \tag{10}$$

$$\text{With} \quad \tilde{x} = T_k(x) \quad \tilde{x}_m = T_k(x_m)$$

where $\max_m \mathrm{sim}(c(x_m), c(x))$ computes the *cosine similarity* between x and its nearest training sample in $\{x_m\}$, f_θ is an auxiliary classifier that aims at determining if x is a shifted example or not, and W_k is the weight vector in the linear layer of $p_{cls}(y^S|x)$. In practice, with M training samples, balancing terms $\lambda_{con}^S = M/\sum_m s_{con}^S$ and $\lambda_{cls}^S = M/\sum_m s_{cls}^S$ are applied to scale the scores of each shifting transformation S. Those balancing terms slightly improve the detection performances, as reported in [53]. Our final anomaly score is $s_{con}(\tilde{x}, \{\tilde{x}_m\}) \cdot \lambda_{con}^S + s_{cls}(\tilde{x}) \cdot \lambda_{cls}^S$.

4　Experiments

4.1　Experiment Setting

We evaluated our methods on six datasets with various imaging protocols (*e.g.* CT, OCT, endoscopy, retinal fundus) and areas (*e.g.* chest, breast, colon, eye). In

Table 1: Semi-supervised fine-grained medical anomaly detection results.

Methods	Extra Training Data	Pnuemonia MNIST	Breast MNIST	SARS-COV-2	Kvasir Polyp	Retinal OCT	APTOS 2019
Reconstruction-based Methods							
GANomaly [1] (ACCV 18)	×	$0.552_{\pm0.01}$	$0.527_{\pm0.01}$	$0.604_{\pm0.00}$	$0.604_{\pm0.00}$	$0.505_{\pm0.00}$	$0.601_{\pm0.01}$
‡UniAD [62] (NeurIPS 22)	×	$0.734_{\pm0.02}$	$0.624_{\pm0.01}$	$0.636_{\pm0.00}$	$0.724_{\pm0.03}$	$0.921_{\pm0.01}$	$0.874_{\pm0.00}$
Normalizing flow-based Methods							
‡CFlow [25] (WACV 22)	×	$0.537_{\pm0.01}$	$0.647_{\pm0.01}$	$0.622_{\pm0.01}$	$0.852_{\pm0.03}$	$0.712_{\pm0.02}$	$0.452_{\pm0.01}$
UFlow [54]	×	$0.792_{\pm0.01}$	$0.631_{\pm0.01}$	$0.653_{\pm0.02}$	$0.562_{\pm0.02}$	$0.630_{\pm0.01}$	$0.731_{\pm0.00}$
FastFlow [63]	×	$0.827_{\pm0.02}$	$0.667_{\pm0.01}$	$0.700_{\pm0.01}$	$0.516_{\pm0.03}$	$0.744_{\pm0.01}$	$0.772_{\pm0.02}$
Teacher-Student Methods							
KDAD [47] (CVPR 21)	×	$0.378_{\pm0.02}$	$0.611_{\pm0.02}$	$0.770_{\pm0.01}$	$0.775_{\pm0.01}$	$0.801_{\pm0.00}$	$0.631_{\pm0.01}$
RD4AD [18] (CVPR 22)	✓	$0.815_{\pm0.01}$	$\mathbf{0.759}_{\pm0.02}$	$0.842_{\pm0.00}$	$0.757_{\pm0.01}$	$\mathbf{0.996}_{\pm0.00}$	$0.921_{\pm0.00}$
†Transformly [17] (CVPR 22)	✓	$0.821_{\pm0.01}$	$0.738_{\pm0.04}$	$0.711_{\pm0.00}$	$0.568_{\pm0.00}$	$0.824_{\pm0.01}$	$0.616_{\pm0.01}$
‡EfficientAD [5] (CVPR 24)	✓	$0.686_{\pm0.02}$	$0.696_{\pm0.03}$	$0.711_{\pm0.02}$	$0.753_{\pm0.03}$	$0.826_{\pm0.02}$	$0.763_{\pm0.02}$
Memory Bank-Based Methods							
CFA [31] (IEEE Access 22)	×	$0.716_{\pm0.01}$	$0.678_{\pm0.02}$	$0.424_{\pm0.03}$	$0.354_{\pm0.01}$	$0.472_{\pm0.01}$	$0.796_{\pm0.01}$
‡PatchCore [44] (CVPR 22)	×	$0.737_{\pm0.01}$	$0.700_{\pm0.02}$	$0.654_{\pm0.01}$	$0.832_{\pm0.01}$	$0.758_{\pm0.01}$	$0.583_{\pm0.01}$
Contrastive Learning-Based Methods							
Meanshift [42] (AAAI 23)	×	$0.818_{\pm0.02}$	$0.648_{\pm0.01}$	$0.767_{\pm0.03}$	$0.694_{\pm0.05}$	$0.438_{\pm0.01}$	$0.826_{\pm0.01}$
CSI [53] Baseline (NeurIPS 20)	×	$0.834_{\pm0.03}$	$0.546_{\pm0.03}$	$0.785_{\pm0.02}$	$0.609_{\pm0.03}$	$0.803_{\pm0.00}$	$0.927_{\pm0.00}$
DIA Ours	×	$\mathbf{0.903}_{\pm0.01}$	$0.750_{\pm0.03}$	$\mathbf{0.851}_{\pm0.03}$	$\mathbf{0.860}_{\pm0.04}$	$0.944_{\pm0.00}$	$\mathbf{0.934}_{\pm0.00}$

†Transformaly is trained under unimodel settings as the original paper.
‡Not support 32×32 resolution, where 128×128 resolution is used for *MNIST datasets.
‡Only 4500 images of the OCT dataset for *PatchCore* are used due to it is the cap for A100.

particular, we experiment on low-resolution datasets of *Pnuemonia MNIST* and *Breast MNIST*, and higher resolution datasets of *SARS-COV-2*, *Kvasir-Polyp*, *Retinal-OCT*, and *APTOS-2019*. A detailed description is in Appendix A.2.

We performed semi-supervised anomaly detection that uses only the normal class for training, namely, the healthy samples. Then we output the anomaly scores for each data instance to evaluate the anomaly detection performance. We use the area under the receiver operating characteristic curve (AUROC) as the metric. All the presented values are computed by averaging at least three runs. Technical details can be found in Appendix A.1. Technically, we use ResNet18 as the backbone model and a batch size of 32. We adopted rotation as the *shifting transformations*, with a fixed $K = 4$ for $0°, 90°, 180°, 270°$. For the *Kvasir-Polyp* dataset, we used perm (*i.e.* jigsaw transformation) since gastrointestinal images are rotation-invariant (details in Appendix B.2). For *dissolving transformations*, all diffusion models are trained on 32×32 images. The diffusion step t is randomly sampled from $t \sim U(100, 200)$ for *Kvasir-Polyp* and $t \sim U(30, 130)$ for the other datasets. For high-resolution datasets, we downsampled images to 32×32 for feature dissolving and then resized them back, avoiding massive computations. Results for different dissolving transformation resolutions are in Section 5.4.

4.2   Results

We compare against 14 previous methods to showcase the performances of our method. Most selected methods are designed for fine-grained anomaly detection

or medical anomaly detection. As shown in Tab. 1, previous work is under-performing or unstable across various fine-grained anomaly detection datasets. Methods that do not leverage external data generally perform less effectively. In contrast, our approach, which employs a fine-grained feature learning strategy, achieves consistently strong and reliable results across all datasets without relying on pretrained models. This highlights the reliability and effectiveness of our strategy, underscoring its superiority in handling diverse medical data modalities and anomaly patterns with stable performances. Notably, our method beats all other methods on four out of six datasets. *RD4AD* takes advantage of pretrained models and achieves better performances on two datasets. In addition, we significantly outperform the baseline *CSI* on all datasets, thereby clearly demonstrating the value of our novel fine-grained feature learning paradigm.

5   Ablation Studies

This section presents a series of ablation studies to understand how our proposed method works under different configurations and parameter settings. In addition, we present results with heuristic blurring methods and shifting transformations in Appendix B, along with the different designs of similarity matrix and non-medical datasets provided in Appendix C.3.

5.1   Dissolving Transformation Steps

We randomly sample dissolving step t from a uniform distribution $U(a, b)$. This experiment investigates various sampling ranges. We establish the minimum step at 30 to ensure minimal changes to the image and assess effectiveness over a 100-step interval. As indicated in Tab. 2, lower steps generally yield better results. The lower step dissolves fine-grained features without significantly altering the coarse-grained image appearance. The model can then focus on the dissolved fine-grained features. Kvasir dataset involves polyps as anomalies, which are pronounced (in the pixel space) compared to the anomalies in other datasets. Consequently, a slightly higher t can lead to enhanced performance.

Table 2: Different diffusion step range.

Step Range	SARS COV-2	Kvasir Polyp	Retinal OCT	APTOS 2019
(30, 130)	**0.851**	0.796	**0.919**	**0.934**
(130, 230)	0.827	**0.860**	0.895	0.920
(230, 330)	0.790	0.775	0.908	0.923
(330, 430)	0.815	0.763	0.896	0.926
(430, 530)	0.803	0.615	0.905	0.926

Table 3: Different training data ratios.

Datasets	DIA ($\gamma = 0.1$)	DIA ($\gamma = 1$)
PneumoniaMNIST	0.745	0.903
Kvasir-Polyp	0.679	0.860

5.2   The Role of Diffusion Models

Given the challenges of acquiring additional medical data, we evaluate how diffusion models affect anomaly detection performances. Specifically, we limit the

training data ratio (γ) for diffusion models to simulate less optimal diffusion models, while keeping other settings unchanged. This experiment examines how anomaly detection performances are impacted when deployed with underperforming diffusion models with insufficient training data. We evaluate on two small datasets where 5856 images are in *PneumoniaMNIST* and 8000 images are in *Kvasir-Polyp*. As shown in Tab. 3, a significant performance drop happened. Thus, better performance of anomaly detection can be obtained with better-trained diffusion models.

A natural next question is, can one utilize well-trained diffusion models to perform dissolving transformations on non-training domains? A well-trained diffusion model is attuned to the attributes of its training dataset. Consequently, it may incorrectly dissolve features if the presented image deviates from the training set. Figure 4 presents the different dissolving effects using diffusion models trained on different datasets. The visual evidence suggests that a data-specific diffusion model accurately dissolves the correct instance-specific features and attempts to revert images towards a more generalized form characteristic of the dataset. In contrast, a diffusion model trained on the CIFAR dataset tends to dissolve the image in a chaotic manner, failing to maintain the image's inherent shape. Additional demonstration with stable diffusion is in Appendix D.

| **(a)** Input | **(b)** $C, t = 200$ | **(c)** $M, t = 200$ | **(d)** $C, t = 400$ | **(e)** $M, t = 400$ |
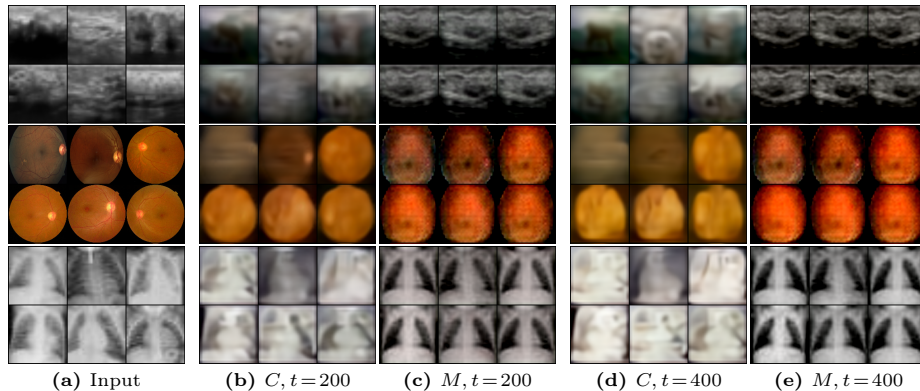
Fig. 4: Dissolving Transformations using different diffusion models. C and M denote if the dissolving transformation is performed based on the diffusion models trained on CIFAR10 or the corresponding dataset, respectively.

5.3   Rotate vs. Perm

Rotate and perm (*i.e.* jigsaw transformation) are reported as the most performant *shifting transformations* [53]. This experiment evaluates their performances under fine-grained settings. As shown in Tab. 4, the rotation transformation outperforms the perm transformation for most datasets. Perm transformation per-

Table 4: Using rotate or perm for shifting transformation.

Method	SARS-COV-2	Kvasir-Polyp	Retinal-OCT	APTOS-2019
DIA-Perm	0.841±0.01	**0.860**±0.01	0.890±0.02	0.926±0.00
DIA-Rotate	**0.851**±0.03	0.813±0.03	**0.944**±0.01	**0.934**±0.00

forms better on the Kvasir dataset since the endoscopic images can be rotation-invariant. In general, the selection of shifting transformations should ease the categorization difficulties associated with the correct shifting distributions. Additional results are in Appendix B.2.

5.4 The Resolution of Feature Dissolved Samples

We use *feature-dissolved* samples with a resolution of 32×32, which significantly improves the anomaly detection performances. Notably, the downsample-upsample routine also dissolves fine-grained features. This experiment investigates the effects of different resolutions for feature-dissolved samples. The experiments adopt 256, 128, 32 batchsizes for the resolution of 32×32, 64×64, 128×128, respectively. As shown in Tabs. 5 and 6, higher resolutions greatly increase computational cost, while it can hardly boost model performances.

The variations in performance across different resolutions are attributed to two main factors. Firstly, the size of training samples impacts this. In larger datasets such as APTOS and Retinal-OCT, the performance degradation is less pronounced. This is because higher-resolution diffusion models require more training data. Secondly, the nature of discriminative features plays a role. High-resolution images naturally contain more details. In datasets like APTOS, where disease indicators are subtler in pixel space (*e.g.* hemorrhages or thinner blood vessels), the performance drop is minimal. In fact, 64x64 resolution images even outperform 32x32 ones for APTOS. Conversely, in datasets like Retinal-OCT, where crucial features are more prominent in pixel space (*e.g.* edemas), lower-resolution images help the model concentrate on these more apparent features. Notably, the computational cost of higher-resolution dissolving transformations is dramatically increased. Our results indicate that a resolution of 32x32 strikes an optimal performance for dissolving effects and computational efficiency.

Table 5: Different resolutions for dissolving transformations.

Dslv. Size	SARS COV-2	Kvasir Polyp	Retinal OCT	APTOS 2019
32	**0.851**	**0.860**	**0.944**	0.934
64	0.803	0.721	0.922	**0.937**
128	0.807	0.730	0.930	0.905

Table 6: Multiply–accumulate operations (MACs) for different resolutions of dissolving transformations. *w/o* denotes no dissolving transformation applied.

Res.	w/o	32×32	64×64	128×128
Params (M)	11.2	19.93	19.93	19.93
MACs (G)	1.82	2.33	3.84	9.90

6   Discussion

Diffusion models work by gradually adding noise to an image over several steps, and then a UNet is employed to learn to reverse this process. During the training of diffusion models, the UNet learns to predict the noise that was added at each step of the diffusion process. This process indirectly teaches the UNet about the underlying structure and characteristics of the data in the dataset. Essentially, the proposed *dissolving transformation* executes a standalone reverse diffusion to reverse the "noise" on non-noisy input images directly. Notably, it still operates under the assumption that there is noise to be removed. Consequently, it interprets the instance-specific fine details and textures of the non-noisy image as noise and attempts to remove them (which we refer to as "dissolve" in our context), as illustrated in Fig. 1. With non-noisy input images from a non-training domain, the diffusion model fails to interpret the correct instance-specific fine details and, therefore, fails to remove the correct features inside the image, as illustrated in Fig. 4. We show additional qualitative results in Appendix D.

Medical image data is particularly suitable for the proposed *dissolving transformations*. Different from other data domains, medical images typically feature a consistent prior, commonly referred to as "atlas" in the medical domain, which is an average representation of a specific patient population, onto which more detailed, instance-specific (discriminative) features are superimposed. For instance, chest X-ray images generally have a gray chest shape on a black background, with additional instance-specific features like bones, tumors, or other pathological findings, being superimposed on top. Those instance-specific features are interpreted by the UNet as "noise" and then removed by the reverse diffusion process. By tuning the hyperparameter t, this process allows for the gradual removal of the most instance-specific features, moving towards the atlas representation of the given image. The feature-dissolved atlas representation serves as a reference for comparison to identify clinically significant changes, while the removed features typically contain pivotal pathological findings. Therefore, to amplify these removed critical features, we deploy a contrastive learning scheme to contrast a given input image and its feature-dissolved counterpart.

7   Conclusion

We proposed an intuitive *dissolving is amplifying* (DIA) method to support fine-grained discriminative feature learning for medical anomaly detection. Specifically, we introduced *dissolving transformations* that can be achieved with a pre-trained diffusion model. We use contrastive learning to enhance the difference between images that have been transformed by dissolving transformations and images that have not. Experiments show *DIA* significantly boosts performance on fine-grained medical anomaly detection without prior knowledge of anomalous features. One limitation is that our method requires training on diffusion models for each of the datasets. In future work, we would like to extend our method to enhance supervised contrastive learning and fine-grained classification by leveraging the fine-grained feature learning strategy.

Acknowledgment

References

1. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: Semi-supervised anomaly detection via adversarial training. In: Computer Vision – ACCV 2018, pp. 622–637. Lecture notes in computer science, Springer International Publishing, Cham (2019)
2. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE (Jul 2019)
3. Angelov, P., Soares, E.: EXPLAINABLE-BY-DESIGN APPROACH FOR COVID-19 CLASSIFICATION VIA CT-SCAN (Apr 2020). `https://doi.org/10.1101/2020.04.24.20078584`, `https://doi.org/10.1101/2020.04.24.20078584`
4. APTOS, A.P.T.O.S.: Aptos 2019 blindness detection. `https://www.kaggle.com/competitions/aptos2019-blindness-detection` (2019)
5. 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)
6. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
7. C. Basilan, M.L.J., https://orcid.org/0000-0003-3105-2252, Padilla, M., https://orchid.org/0000-0001-5025-12872, maleticiajose.basilan@deped.gov.ph, maycee.padilla@deped.gov.ph, Department of Education- SDO Batangas Province, Batangas, Philippines: Assessment of teaching english language skills: Input to digitized activities for campus journalism advisers. International Multidisciplinary Research Journal **4**(4) (Jan 2023)
8. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS) (2020)
9. Chalapathy, R., Chawla, S.: Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407 (2019)
10. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
11. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.: Big self-supervised models are strong semi-supervised learners. arXiv preprint arXiv:2006.10029 (2020)
12. Chen, X., Fan, H., Girshick, R., He, K.: Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297 (2020)
13. Chen, X., He, K.: Exploring simple siamese representation learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (Jun 2021). `https://doi.org/10.1109/cvpr46437.2021.01549`, `https://doi.org/10.1109/cvpr46437.2021.01549`
14. Chen, Y., Zhou, X.S., Huang, T.S.: One-class svm for learning in image retrieval. In: Proceedings 2001 international conference on image processing (Cat. No. 01CH37205). vol. 1, pp. 34–37. IEEE (2001)

15. Cheng, H., Liu, H., Gao, F., Chen, Z.: ADGAN: A scalable GAN-based architecture for image anomaly detection. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE (Jun 2020). https://doi.org/10.1109/itnec48623.2020.9085163

16. Cho, H., Seol, J., goo Lee, S.: Masked contrastive learning for anomaly detection. In: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization (Aug 2021). https://doi.org/10.24963/ijcai.2021/198, https://doi.org/10.24963/ijcai.2021/198

17. Cohen, M.J., Avidan, S.: Transformaly - two (feature spaces) are better than one. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 4060–4069 (June 2022)

18. 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 (CVPR). pp. 9737–9746 (June 2022)

19. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1. p. 766–774. NIPS'14, MIT Press, Cambridge, MA, USA (2014)

20. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)

21. Golan, I., El-Yaniv, R.: Deep anomaly detection using geometric transformations. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS) (2018)

22. Goldstein, M., Dengel, A.: Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. KI-2012: poster and demo track **1**, 59–63 (2012)

23. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014)

24. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P.H., Buchatskaya, E., Doersch, C., Pires, B.A., Guo, Z.D., Azar, M.G., Piot, B., Kavukcuoglu, K., Munos, R., Valko, M.: Bootstrap your own latent: A new approach to self-supervised learning (2020)

25. 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)

26. Gutmann, M.U., Hyvärinen, A.: Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. Journal of machine learning research **13**(2) (2012)

27. Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., Nakayama, H.: Gan-based synthetic brain mr image generation. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). pp. 734–738. IEEE (2018)

28. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. arXiv preprint arXiv:1911.05722 (2019)

29. Ker, J., Wang, L., Rao, J., Lim, T.: Deep learning applications in medical image analysis. Ieee Access **6**, 9375–9389 (2017)

30. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
31. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access **10**, 78446–78454 (2022)
32. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
33. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B.: Adversarial autoencoders. In: International Conference on Learning Representations (2016), https://arxiv.org/abs/1511.05644v2
34. Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., et al.: Mixed precision training. arXiv preprint arXiv:1710.03740 (2017)
35. Murase, H., Fukumizu, K.: Algan: Anomaly detection by generating pseudo anomalous data via latent variables. IEEE Access **10**, 44259–44270 (2022). https://doi.org/10.1109/ACCESS.2022.3169594
36. Musa, T.H.A., Bouras, A.: Anomaly detection: A survey. In: Proceedings of Sixth International Congress on Information and Communication Technology, pp. 391–401. Springer Singapore (Oct 2021). https://doi.org/10.1007/978-981-16-2102-4_36, https://doi.org/10.1007/978-981-16-2102-4_36
37. Pang, G., Shen, C., Cao, L., Van Den Hengel, A.: Deep learning for anomaly detection. ACM Comput. Surv. **54**(2), 1–38 (Mar 2021)
38. Pang, G., Shen, C., Jin, H., van den Hengel, A.: Deep weakly-supervised anomaly detection (2019)
39. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference (MMSYS). pp. 164–169 (2017). https://doi.org/10.1145/3083187.3083212
40. Pourreza, M., Mohammadi, B., Khaki, M., Bouindour, S., Snoussi, H., Sabokrou, M.: G2d: generate to detect anomaly. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2003–2012 (2021)
41. Rani, B.J.B., E, L.S.M.: Survey on applying GAN for anomaly detection. In: 2020 International Conference on Computer Communication and Informatics (ICCCI). IEEE (Jan 2020). https://doi.org/10.1109/iccci48352.2020.9104046
42. Reiss, T., Hoshen, Y.: Mean-shifted contrastive loss for anomaly detection. arXiv preprint arXiv:2106.03844 (2021)
43. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2021)
44. 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)
45. Ruff, L., Vandermeulen, R.A., Görnitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: Proceedings of the 35th International Conference on Machine Learning. vol. 80, pp. 4393–4402 (2018)
46. Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3379–3388 (2018)
47. 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)

48. Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., Barfett, J.: Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 990–994. IEEE (2018)

49. Salem, M., Taheri, S., Yuan, J.S.: Anomaly generation using generative adversarial networks in host-based intrusion detection. In: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). pp. 683–687. IEEE (2018)

50. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Lecture Notes in Computer Science, pp. 146–157. Lecture notes in computer science, Springer International Publishing, Cham (2017)

51. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery (2017)

52. Shekarizadeh, S., Rastgoo, R., Al-Kuwari, S., Sabokrou, M.: Deep-disaster: Unsupervised disaster detection and localization using visual data (2022)

53. Tack, J., Mo, S., Jeong, J., Shin, J.: Csi: Novelty detection via contrastive learning on distributionally shifted instances. Advances in neural information processing systems **33**, 11839–11852 (2020)

54. Tailanian, M., Pardo, Á., Musé, P.: U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold. arXiv preprint arXiv:2211.12353 (2022)

55. Thudumu, S., Branch, P., Jin, J., Singh, J.J.: A comprehensive survey of anomaly detection techniques for high dimensional big data. Journal of Big Data **7**(1), 1–30 (2020)

56. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: European conference on computer vision. pp. 499–515. Springer (2016)

57. Williams, G., Baxter, R., He, H., Hawkins, S., Gu, L.: A comparative study of rnn for outlier detection in data mining. In: 2002 IEEE International Conference on Data Mining, 2002. Proceedings. pp. 709–712. IEEE (2002)

58. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 650–656 (June 2022)

59. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification. arXiv preprint arXiv:2110.14795 (2021)

60. Ye, F., Huang, C., Cao, J., Li, M., Zhang, Y., Lu, C.: Attribute restoration framework for anomaly detection. IEEE Transactions on Multimedia **24**, 116–127 (2022). `https://doi.org/10.1109/tmm.2020.3046884`

61. You, Y., Gitman, I., Ginsburg, B.: Large batch training of convolutional networks. arXiv preprint arXiv:1708.03888 (2017)

62. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS) (06 2022)

63. 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)
64. Zhang, Y., Ling, H., Gao, J., Yin, K., Lafleche, J.F., Barriuso, A., Torralba, A., Fidler, S.: Datasetgan: Efficient labeled data factory with minimal human effort. In: CVPR (2021)
65. Zhao, Z., Li, B., Dong, R., Zhao, P.: A surface defect detection method based on positive samples. In: Lecture Notes in Computer Science, pp. 473–481. Springer International Publishing (2018). `https://doi.org/10.1007/978-3-319-97310-4_54`