Anomaly Detection using Score-based Perturbation Resilience

Woosang Shin[* 1,2]   Jonghyeon Lee[* 1,2]   Taehan Lee[1]   Sangmoon Lee[2]   Jong Pil Yun[† 1,3]

[1]Korea Institute of Industrial Technology (KITECH), South Korea
[2]Kyungpook National University (KNU), South Korea
[3]University of Science and Technology (UST), South Korea

{welldone, ljh0325, taehanlee}@kitech.re.kr   moony@knu.ac.kr   rebirth@kitech.re.kr

Abstract

Unsupervised anomaly detection is widely studied in industrial applications where anomalous data is difficult to obtain. In particular, reconstruction-based anomaly detection can be a feasible solution if there is no option to use external knowledge, such as extra datasets or pre-trained models. However, reconstruction-based methods have limited utility due to poor detection performance. A score-based model, also known as a denoising diffusion model, recently has shown a high sample quality in the generation task. In this paper, we propose a novel unsupervised anomaly detection method leveraging the score-based model. The proposed method shows promising performance without requiring external knowledge. The score, a gradient of the log-likelihood, has a property that is available for anomaly detection. The samples on the data manifold can be restored instantly by the score, even if they are randomly perturbed. We call this score-based perturbation resilience. On the other hand, the samples that deviate from the manifold cannot be restored in the same way. The variation of resilience depending on the sample position can be an indicator to discriminate anomalies. We derive this statement from a geometric perspective. Our method shows superior performance on three benchmark datasets for industrial anomaly detection. Specifically, on MVTec AD, we achieve image-level AUROC of 97.7% and pixel-level AUROC of 97.4% outperforming previous works that do not use external knowledge.

1. Introduction

Anomaly detection is an important task in the manufacturing industry. It finds its usage in production lines, such as quality control through defect inspection. Consistent qual-
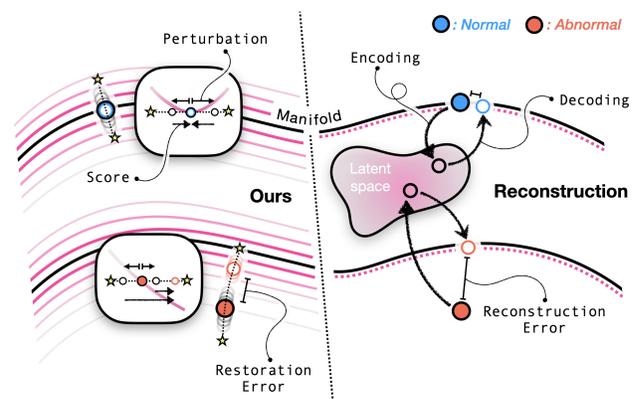
---

[*]Equal Contribution
[†]Corresponding author

Figure 1. Illustrations of the proposed method (**Left**) and the reconstruction-based method (**Right**) for unsupervised anomaly detection. The pink gradation solid line and dotted line denote the negative log-likelihood and the manifold estimated by the reconstruction model, respectively.

ity control can be ensured by an automated anomaly detection system which reduces human labor and prevents the decrease in attention spans. Supervised anomaly detection is a simple approach. However, sometimes, the supervised approach is not feasible in practical applications due to various reasons. On real-world manufacturing lines, abnormal samples are rare. As a result, only a few abnormal samples are available. Because abnormal samples are irregular and have a large diversity, identifying abnormal patterns with a small amount of data is difficult. Moreover, obtaining accurate ground truth is time-consuming and requires experts with domain knowledge. Thus, gathering an adequate dataset for training the supervised method is impractical. Therefore, an unsupervised approach that detects abnormal samples using only abundant normal samples is widely used. The unsupervised approach to anomaly detection encourages learning useful features from only normal samples, with no prior

knowledge of abnormality. This approach is used not only in the manufacturing industry but also in medical diagnostics [53, 30, 10], video surveillance [36, 21, 13], and autonomous driving [18, 44, 15].

Recently, various unsupervised approaches have been introduced. Embedding-based methods and synthetic anomaly-based methods utilize pre-trained models with large datasets, such as ImageNet, and artificially synthesized anomalous data, respectively, instead of specific supervision. These methods have shown remarkable detection performance in industrial benchmarks. However, despite their superior performance, a limitation remains in that the transferability of leveraged external knowledge needs to be verified with some degree of priori. The reconstruction-based approach can be another good alternative. The right side of Figure 1 schematically shows this approach from a geometrical perspective. The reconstruction model, such as a variational auto-encoder, is trained to mimic the mapping function from latent space, also known as feature space, to the data manifold. A difference between the original sample and the sample reconstructed by a reconstruction model, denoted by reconstruction error, is used as a discriminative measure for anomaly detection. The latent space of the model can reflect only the representation of data used to train. Therefore, for unseen anomalous samples, the reconstruction error will be larger than for normal samples. Not only is this approach theoretically intuitive and reasonable, but it is also practically versatile in that it requires little external knowledge.

We propose an unsupervised anomaly detection method using a score-based model. Score-based models have attracted significant attention in the field of data generation because of their capability to produce high-quality samples. There are previous studies utilizing this model for anomaly detection in the medical imaging field [47, 45]. They have shown impressive results. However, the denoising diffusion process involved in their proposed method is time-consuming, making it difficult to apply to manufacturing industries that require real-time processing. Instead, we have found an inherent property of the score, a gradient of the log-likelihood, that is available to detect anomalies. The score can immediately restore the samples on the data manifold, even if perturbed randomly. We call this *score-based resilience to perturbations*. On the other hand, it is impossible for the samples that deviate from the manifold to be restored through the score. The left side of Figure 1 shows this principle. The variation of resilience depending on the sample position can be a discriminator for anomaly detection. We introduce the stochastic measure, restoration error, to quantify the resilience of the observed sample. The restoration error is defined as an expectation of the distance between the restored sample after being perturbed and the original sample.

Our key contributions are summarized as follows:

- We propose a novel approach leveraging a score-based model for unsupervised anomaly detection. The proposed method requires no external knowledge, such as pre-trained models or extra datasets.

- Interpreting a score function from a geometric view, we support the theoretical reasonability that score-based resilience to perturbation can be available for anomaly detection. In addition, we introduce a stochastic measure, restoration error, to quantify resilience.

- Conducting experiments on public benchmark datasets, MVtec AD, BTAD, and MPDD, for industrial anomaly detection, we show that our method can achieve outperforming results in an unsupervised setting.

2. Related Works

2.1. Reconstruction-based Methods

The core assumption of reconstruction-based methods is that a model trained only on normal samples cannot accurately reconstruct an abnormal region. They use auto-encoders [9, 37], variational auto-encoders [7, 19], and generative adversarial networks [28, 1, 2] for image reconstruction. The error between the input image and the reconstructed image is used to calculate the anomaly score. However, it often fails to detect abnormal regions, due to a well-generalization of neural networks. RIAD [50] and InTra [23] proposed integration with the inpainting task to avoid generalizing to abnormal regions. The MemAE was proposed in [9], MemAE introduced memory modules to avoid the generalization problem. VAE-grad [7] proposed an approach using gradient descent to the energy derived from the loss function of auto-encoders. TrustMAE [37] is an improvement to MemAE. CAVGA [40] identifies anomalies as convolutional latent variables to preserve spatial information. MetaFormer[46] uses meta-learning to tackle the issues of model adaptation and reconstruction gap.

2.2. Embedding-based Methods

In embedding-based anomaly detection, the anomaly score is based on the distance between the embedding vectors of the normal sample and the test sample. The difference between embedding vectors is measured using Euclidean distance, Mahalanobis distance, Gaussian distribution, KNN, etc. Deep-SVDD [27] finds the smallest hypersphere that wraps around the normal data and identifies anomalies based on Euclidean distance between its centroid and the embedding vectors of test samples. Patch-SVDD [48] extended Deep-SVDD to detect anomalies at
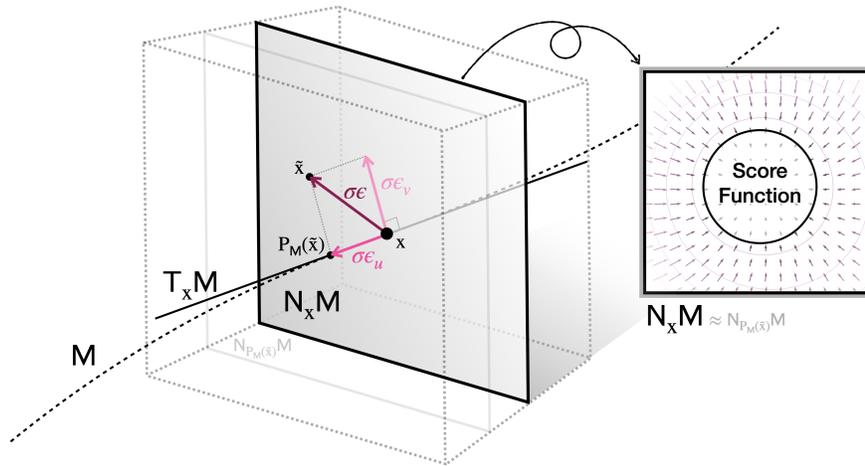
Figure 2. Overview of the geometric interpretation. The cube with a gray dotted outline denotes a neighborhood of the point x on the manifold M satisfied the linearity. Point $\tilde{x}$ is a perturbed point by a random vector $\sigma\epsilon$ from x. Vector $\sigma\epsilon_v$ and $\sigma\epsilon_u$ denote the perpendicular and tangential element of $\sigma\epsilon$ with respect to M, respectively. And, $P_M(\tilde{x})$ denotes the projected point from $\tilde{x}$ onto M. The gray box with a black border denotes a normal space $N_x M \perp T_x M$, where $T_x M$ denotes a tangential space. The vector field contained in the box with a black and gray border on the right side represents a score function on $N_x M$.

the patch level. Recent works [5, 6, 25] take advantage of the powerful feature extraction of pre-trained models. SPADE [5] utilizes hierarchical features extracted from pre-trained models and pixel-wise k-Nearest Neighbors algorithm for anomaly detection. PaDiM [6] uses features from several layers of a pre-trained model and multivariate Gaussian distributions to model the distribution of normal samples. The Mahalanobis distance is then used as an anomaly score. SPADE and PaDiM are known as memory bank-based methods. These methods require a complex feature matching process for inference. To reduce the complexity of inference, PatchCore [25] introduces greedy coreset sub-sampling [31] to reduce redundancy in patch-level memory banks.

2.3. Synthetic Anomaly-based Methods

Recently, methods based on simulated synthetic anomalies have been proposed. For example, CutPaste [17] creates synthetic abnormal samples by pasting a cropped image patch onto a normal image and learns the representation by classifying the normal sample. In DRÆM [49], pseudo anomaly regions are created by combining an additional textured image with a random noise mask. The reconstructive sub-network, then, reconstructs the pseudo anomaly image, and the discriminative sub-network compares the input image with the reconstructed image to determine the anomaly. In [29], NSA extended CutPaste [17] approach. NSA integrates Poisson image editing to create a more natural synthetic abnormal sample.

3. Background

3.1. Data Manifold Hypothesis

The hypothesis that data distribution only has support on low dimensional structured space, data manifold, embedded in high dimensional space, ambient space, has become commonly accepted in the unsupervised setting. Many algorithms linearly approximating the local region of the underlying data manifold have developed, such as Locally Linear Embedding [26], Isomap [39] and Manifold Parzen Windows [42]. In addition, some algorithms, such as denoising auto-encoder [43] and Non-Local Manifold Tangent Learning [3], aim to capture the representation of the data manifold. The studies [32, 34, 11] that have motivated us are also related to this hypothesis. The manifold has a formal definition, mathematically, but we exploit only some geometric properties related to the data manifold.

Suppose that the data X is uniformly distributed on the manifold $M \subset \mathbb{R}^d = \text{supp}(X)$, and M is a globally smooth structure. Recall the proposition of unsupervised anomaly detection that we can deal with only the normal data. Namely, M denotes the normal manifold. Then, there exists a vector space of dimensionality $l \ll d$, which contains all of the tangent vectors to M at point $x \in M$. The vector space can be called tangent space and denoted by $T_x M$. The tangent space is almost the same as a linear approximation to M at the neighborhood of x. In addition, we state the set of perpendicular vectors to tangent space at x as normal space $N_x M \cong \mathbb{R}^{d-l}$.

3.2. Score Matching

In unsupervised learning literature, there are a number of approaches to estimate the density of the data distribution $p(\mathrm{x})$, or log-likelihood, $\log p(\mathrm{x})$. Score matching was introduced by Hyvärinen to circumvent the computation of intractable partition function terms contained in parametric modes often used in machine learning to estimate data density [12]. Vincent established the connection between the denoising auto-encoder and the score-matching by proving the equivalence of the denoising auto-encoder objective and the score-matching [41]. We use this objective, also known as denoising score-matching, to learn an approximation model of the score function.

4. Proposed Method

4.1. Interpretation of Score in a Geometric View

We analyze the connection between a score function and perturbation from geometric and probabilistic perspectives. Let $\tilde{\mathrm{x}}$ be a perturbed point by random vector $\sigma\epsilon$. The probability distribution of $\tilde{\mathrm{x}}$ is

$$p(\tilde{\mathrm{x}}) = \int p(\tilde{\mathrm{x}}|\mathrm{x})p(\mathrm{x})\mathrm{dx}, \qquad (1)$$

where $p(\tilde{\mathrm{x}}|\mathrm{x})$ is a perturbation kernel. If we apply a gaussian kernel $\mathrm{N}(\tilde{\mathrm{x}}; \mathrm{x}, \sigma^2 I)$ to perturbation kernel, the score, $\nabla_{\tilde{\mathrm{x}}} \log p(\tilde{\mathrm{x}})$, can be derived as

$$\nabla_{\tilde{\mathrm{x}}} \log p(\tilde{\mathrm{x}}) = \frac{1}{\sigma^2}\big(\mathrm{E}[\mathrm{x}|\tilde{\mathrm{x}}] - \tilde{\mathrm{x}}\big), \qquad (2)$$

where $\mathrm{E}[\mathrm{x}|\tilde{\mathrm{x}}] = \int \mathrm{x}p(\mathrm{x}|\tilde{\mathrm{x}})\mathrm{dx}$ for x given $\tilde{\mathrm{x}}$. From the data manifold hypothesis, we assume that the data is uniformly distributed on the manifold and the tangent space locally coincides with the manifold. Holding these assumptions, the conditional expectation is

$$\mathrm{E}[\mathrm{x}|\tilde{\mathrm{x}}] = \mathrm{P_M}(\tilde{\mathrm{x}}), \qquad (3)$$

where $\mathrm{P_M}(\tilde{\mathrm{x}})$ denotes orthogonally projected point onto the manifold from $\tilde{\mathrm{x}}$. This fact implies that a score function of $\tilde{\mathrm{x}}$ is an orthogonal projector. We provide the detailed derivation of Equation 2 and Equation 3 in the supplementary material.

4.2. Perturbation Resilience based on Score

Let the sample be resilient to perturbation if it, after being perturbed, can be restored to its original state by the score. In the previous section, we have shown that the score function is the locally orthogonal projector. Let us set $\tilde{\mathrm{x}} = \mathrm{x} + \sigma\epsilon$, where ϵ denotes the random perturbation and $\mathrm{x} \in \mathrm{M}$ (i.e. normal sample). Considering the local linearity of M, $\mathrm{N_{P_M(\tilde{x})}M} \simeq \mathrm{N_x M}$. Then, $\sigma\nabla_{\tilde{\mathrm{x}}} \log p(\tilde{\mathrm{x}})$ become
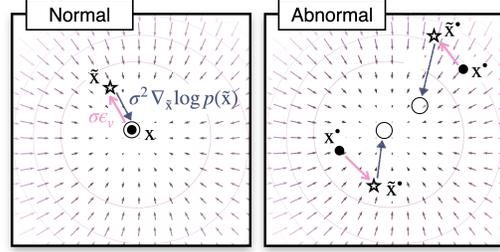

Figure 3. Comparison of the perturbation resilience based on the score function for normal and abnormal samples. The directional arrows that denote the vectors show the relations between the perpendicular element of the random perturbation ϵ_v and the score $\sigma\nabla_{\tilde{\mathrm{x}}} \log p(\tilde{\mathrm{x}})$ in the normal space $\mathrm{N_x M}$. And, $\mathrm{x}^\bullet$ denotes the deviated sample from M (i.e. abnormal sample).

a counterpart of the perpendicular element ϵ_v of ϵ with respect to M. Thus, the distance between the original point and the restored point is $\|\mathrm{x} - \mathrm{P_M}(\tilde{\mathrm{x}})\|^2 = \sigma^2\|\epsilon_u\|^2$, where ϵ_u denotes the tangential element of ϵ with respect to M and $\epsilon_u \perp \epsilon_v$. For small enough σ, $\sigma^2\|\epsilon_u\|^2$ can be ignored. The score function, consequently, can restore the perturbed sample $\tilde{\mathrm{x}}$ to the original state x, as shown on the left in Figure 3, if $\mathrm{x} \in \mathrm{M}$. In other words, this sample is resilient to perturbation according to our definition. On the other hand, we can easily infer that the resilience of $\mathrm{x}^\bullet$ is even lower than that of x through the intuition, as shown on the right in Figure 3, that the score at the perturbed sample $\tilde{\mathrm{x}}^\bullet$ is unable to compensate for ϵ_v.

4.3. Score Function Approximation

In high-dimensional space, the perturbation by random vector (e.g., Gaussian noise) has a gap from our intuition. As dimensionality d increases, the isotropic gaussian distribution with variance scale σ^2 becomes rapidly close to the uniform distribution on the surface of a ball of radius $\sigma\sqrt{d}$. This phenomenon is also known as the Gaussian Annulus Theorem. For this reason, the distribution of perturbed samples by single scale variance would be concentrated at a specific level. In order to avoid the approximation model being trained only for a specific level, we borrow the perturbation strategy with multi-scale variance from [34, 11, 33].

Let $\tilde{\mathrm{x}}_k$ denote the noisy data by additive gaussian noise with multi-scale variance σ_k^2, where $\sigma_k = \sqrt{(\sigma^{2k} - 1)/2\log\sigma}$ for $k \in (0, 1]$. Then, the denoising score-matching objective is defined as

$$L(\theta) = \mathrm{E}_{p_X(\mathrm{x})}\mathrm{E}_{p(\tilde{\mathrm{x}}_k|\mathrm{x})}[\|s_\theta(\tilde{\mathrm{x}}_k, k) - (\frac{\mathrm{x} - \tilde{\mathrm{x}}_k}{\sigma_k})\|^2], \quad (4)$$

where $s_\theta(\tilde{\mathrm{x}}_k, k)$ and $p_X(\mathrm{x})$ denote the approximation model conditioned with k and the probability distribution

of the normal data, respectively. This formula allows the optimized score model $s_{\theta*}(\tilde{x}_k; k) \approx \sigma_k \nabla_{\tilde{x}_k} \log p(\tilde{x}_k)$.

4.4. Restoration Error Estimation

We have shown that score-based perturbation resilience can serve as a discriminative measure for anomaly detection. Therefore, for the quantification of the resilience of the sample to perturbation, we introduce the reconstruction error. Let $\hat{x}_\theta(x; k) = \tilde{x}_k + \sigma_k s_\theta(\tilde{x}_k; k)$ be the restored point by $s_\theta(\tilde{x}_k; k)$, where $\tilde{x}_k = x + \sigma_k \epsilon$ and $\epsilon \sim N(0, I)$. The variable k for $\hat{x}_\theta(x; k)$ is constant in the inference phase. Then, the restoration error of x is defined as

$$\mathrm{E}\big[\|x - \hat{x}_\theta(x; k)\|^2\big] = \sigma_k^2 \, \mathrm{E}\big[\|\epsilon - s_\theta(\tilde{x}_k; k)\|^2\big], \quad (5)$$

with respect to $s_\theta(\tilde{x}_k; k)$. In practice, to localize the anomalous region in x, we introduce the reconstruction error map. The reconstruction error map $\mathbf{e}_R(x)$ can be computed by

$$\mathbf{e}_R(x)_{ij} = \mathrm{E}\big[\|\epsilon_{ij} - s_\theta(\tilde{x}_k; k)_{ij}\|^2\big], \quad (6)$$

where i and j are the pixel index of an image, and $\mathbf{e}_R(x)_{ij}$ is (i, j)th elements of $\mathbf{e}_R(x)$. Since k is a constant, σ_k^2 can be omitted. In addition, we apply a spatial smoothing operation to $\mathbf{e}_R(x)$. We use this as a discriminative measure for pixel-level anomaly detection (i.e., anomaly localization). For image-level anomaly detection, $\max \mathbf{e}_R(x)$ is used. This metric is also, in general, referred to as anomaly score. Note that this "*score*" does not mean the same as the score mentioned in this paper.

The description of the proposed method is summarized in Algorithm 1. Although estimating the restoration error involves multiple prediction as stated in Algorithm 1, it can be efficiently resolved through parallel processing because it is not a sequential process. Note that parallel processing can improve inference speed.

Algorithm 1 Restoration Error Estimation
___

Input: Optimized score model $S_{\theta*}$, Target sample x, Num predictions N, Variance scale parameters (σ, k), Smoothing filter $\mathcal{K}$

Output: Restoration error map $\mathbf{e}_R(x)$

    Initialisation :
1:   $h, w, c \leftarrow$ height, width, channel of x
2:   $\sigma_k \leftarrow \sqrt{(\sigma^{2k} - 1)/2\log\sigma}$
3:   $\mathbf{e}_R(x) \leftarrow 0 \in \mathbb{R}^{(h \times w)}$
4:   **for** $n = 1$ to N **do**
5:     $\epsilon \sim N(0, I) \in \mathbb{R}^{(h \times w \times c)}$
6:     $\tilde{x}_k \leftarrow x + \sigma_k \epsilon$
7:     $\mathbf{e}_R(x)_{ij} \leftarrow \mathbf{e}_R(x)_{ij} + \|\epsilon_{ij} - s_\theta^*(\tilde{x}_k; k)_{ij}\|^2$
8:   **end for**
9:   $\mathbf{e}_R(x) \leftarrow \mathbf{e}_R(x) * \mathcal{K}$ {$*$: spatial-wise convolution}
10:   **return** $\mathbf{e}_R(x)$
___

5. Experiments

5.1. Datasets

We conduct experiments on a variety of industrial anomaly detection datasets (MVTec AD [4], Metal Parts Defect Detection (MPDD) [14], and BeanTech Anomaly Detection (BTAD) [22]) to demonstrate the effectiveness of the proposed method.

MVTec AD   MVTec AD contains sub-datasets of 15 categories (textures of 10 and objects of 5). There are 60 to 320 high-resolution images for each category. The image resolutions are in the range between 700×700 and 1024×1024 pixels. The training set has a total of 3,629 defect-free images. And, the test set has defect-free images and defect images of 1,725. The defect types in the test set include cracks, scratches, broken, and holes. Pixel-level ground truth masks are also included for each defect image in order to evaluate anomaly localization performance.

MPDD   MPDD is introduced for defect detection of painted metal parts in 6 categories. The images were collected under various conditions, such as spatial orientations, distance between objects, position of objects, light luminance and backgrounds, taking into account the actual manufacturing process. The training set contains 888 defect-free images and the test set contains 176 defect-free images and 282 defect images. The resolution is 1024×1024 pixels.

BTAD   BTAD contains a total of 2830 images for three product categories. Product 01, Product 02, and Product 03 have 400, 1000, and 399 train images respectively. The resolution of the three product categories is 1600×1600 pixels, 600×600 pixels, and 800×600 pixels, respectively.

5.2. Evaluation Metrics

The area under the receiver operating characteristic (AU-ROC) is a standard metric for anomaly detection. As in previous studies [4, 29, 25], we employ image-level AU-ROC and pixel-level AUROC to measure the performance of anomaly detection and localization, respectively. We report both individual AUROC for each category and average AUROC.

5.3. Implementation Details

For experiments, the network architecture and hyper-parameters are taken almost directly from previous studies[24, 35]. We use a U-net-based network architecture with the sinusoidal positional embedding for conditioning on time steps used in [35]. A detailed description of the

Method	Average	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule	HZ	MN	Pill	Screw	TB	TS	Zipper
						With external knowledge										
DRÆM[49]	97.3	95.5	99.7	98.6	99.2	96.4	99.1	94.7	94.3	99.7	99.5	97.6	97.6	98.1	90.9	98.8
RD [8]	97.8	98.9	99.3	99.4	95.6	95.3	98.7	97.4	98.7	98.9	97.3	98.2	99.6	99.1	92.5	98.2
PatchCore [25]	98.1	99.0	98.7	99.3	95.6	95.0	98.6	98.4	98.8	98.7	98.4	97.4	99.4	98.7	96.3	98.8
						Without external knowledge										
AE(MSE)* [7]	80.4	53.9	96.0	75.1	47.6	63.0	90.9	73.2	78.6	97.6	88.0	88.5	97.9	97.1	90.6	68.0
AE(SSIM)* [7]	81.8	54.5	96.0	71.0	49.6	64.1	93.3	79.0	76.9	96.6	88.1	89.5	98.3	97.3	90.4	82.8
γ-VAE grad.* [7]	88.8	72.7	97.9	89.7	58.1	80.9	93.1	88.0	91.7	98.8	91.4	93.5	97.2	98.3	93.1	87.1
TrustMAE* [37]	93.9	97.4	99.1	95.1	97.3	99.8	97.0	85.1	78.8	98.5	76.1	83.3	82.4	96.9	87.5	87.5
RIAD* [50]	94.2	96.3	98.8	99.4	89.1	85.8	98.4	84.2	92.8	96.1	92.5	95.7	98.8	98.9	87.7	97.8
Patch-SVDD [48]	95.7	92.6	96.2	97.4	91.4	90.8	96.8	98.1	95.8	97.5	98.0	95.1	95.7	98.1	97.0	95.1
CutPaste [17]	96.0	98.3	97.5	99.5	90.5	95.5	97.6	90.0	97.4	97.3	93.1	95.7	96.7	98.1	93.0	99.3
NSA [29]	96.3	95.5	99.2	99.5	99.3	90.7	98.3	96.0	97.6	97.6	98.4	98.5	96.5	94.9	88.0	94.2
Ours	97.4	96.4	98.9	99.3	96.8	95.4	95.9	96.9	96.6	98.7	96.6	98.2	99.5	97.8	94.7	98.8

Table 1. Performance comparison of anomaly localization in pixel-level AUROC(%) on MVTec AD. For clarity, the category names Hazelnut, Metal Nut, Toothbrush, and Transistor are abbreviated as HN, MN, TB, and TS respectively. *: reconstruction-based method

network structure can be found in the supplementary material. For the training phase, k is randomly sampled from a uniform distribution on $(0, 1]$. For the inference phase, the perturbation scale parameters (σ, k) are set to (25, 0.001). The number of predictions N is set to 3. We train the model from scratch with a batch size of 16 for 3000 epochs. Image augmentation is not used because it requires prior knowledge. For MVTec AD [4], the network is trained using the AdamW [20] optimizer with $\beta = (0.9, 0.999)$ and a learning rate of 0.0001 which is reduced by a factor of 10 at epoch 1000 and 2000. For MPDD [14] and BTAD [22], the learning rate is set to 0.0001 without any learning rate schedule. The weight decay is set to 0.001. The exponential moving average (EMA) technique is used and the EMA rate is 0.9999. We resize the images from all datasets to 192 × 192 pixels. For a fair comparison with previous works, the anomaly maps are resized to 256 × 256 pixels. We resize the images and the anomaly maps using Lanczos interpolation. The anomaly maps are smoothed by a 31 × 31 averaging filter. Our method is implemented by PyTorch 1.11.0 and CUDA 11.1, and all experiments run with NVIDIA RTX A6000 GPU. All experimental settings except the learning rate scheduler are identical regardless of the datasets.

5.4. Baseline Methods

MVTec AD For anomaly localization, we compare our method with AE(MSE) [7], AE(SSIM) [7], γ-VAE grad. [7], TrustMAE [37], RIAD [50], CutPaste [17], Patch-SVDD [48], NSA [29], DRÆM [49], RD [8], and Patch-Core [25]. The results for AE(MSE) and AE(SSIM) are reported in [7]. For anomaly detection, we compare our methods with Metaformer [46], DRÆM [49], DSR [51], Patch-Core [25], TrustMAE [37], RIAD [50], Patch-SVDD [48], CutPaste [17], and NSA [29].

Method	Average	Product 01	Product 02	Product 03
		With external knowledge		
FYD [52]	97.0	96.1	95.3	99.7
		Without external knowledge		
AE(MSE)* [22]	78.0	49.0	92.0	95.0
AE(MSE+SSIM)* [22]	79.0	53.0	96.0	89.0
VT-ADL* [22]	90.0	99.0	94.0	77.0
Patch-SVDD [48]	93.1	94.9	92.7	91.7
Ours	96.0	96.7	96.2	95.2

Table 2. Performance comparison of anomaly localization in pixel-level AUROC(%) on BTAD. *: reconstruction-based method

Method	Average	BBK	BBN	BWH	CN	MP	Tubes
		With external knowledge					
SemiOrthogonal [16]	88.2	85.3	89.6	94.1	94.6	85.9	79.9
PatchCore [25]	95.7	98.4	91.5	97.4	95.0	96.6	95.1
SPADE [5]	95.9	94.3	97.2	96.8	98.4	93.0	95.9
PaDiM [6]	96.7	94.2	92.4	98.1	97.9	92.9	93.9
		Without external knowledge					
Skip-GANomaly* [2]	82.2	89.0	87.1	78.8	80.2	89.7	77.3
DAGAN* [38]	83.3	89.7	81.5	70.6	85.7	90.0	82.3
Ours.	96.8	99.0	97.7	86.2	99.4	99.0	99.4

Table 3. Performance comparison of anomaly localization in pixel-level AUROC(%) on MPDD. For clarity, the category names Bracket black, Bracket brown, bracket white, Connector, Metal plate are abbreviated as BBK, BBN, BWH, CN and MP respectively. *: reconstruction-based method

BTAD The comparison baselines include AE(MSE), AE(SSIM), VT-ADL [22], Patch-SVDD [48], and FYD [52]. The results for AE(MSE) and AE(SSIM) are reported in [22].

MPDD We consider methods including Skip-GANomaly [2], DAGAN [38], SPADE [5], SemiOrthogonal [16], PaDiM [6], and PatchCore [25]. The results of these methods are all reported in [14].
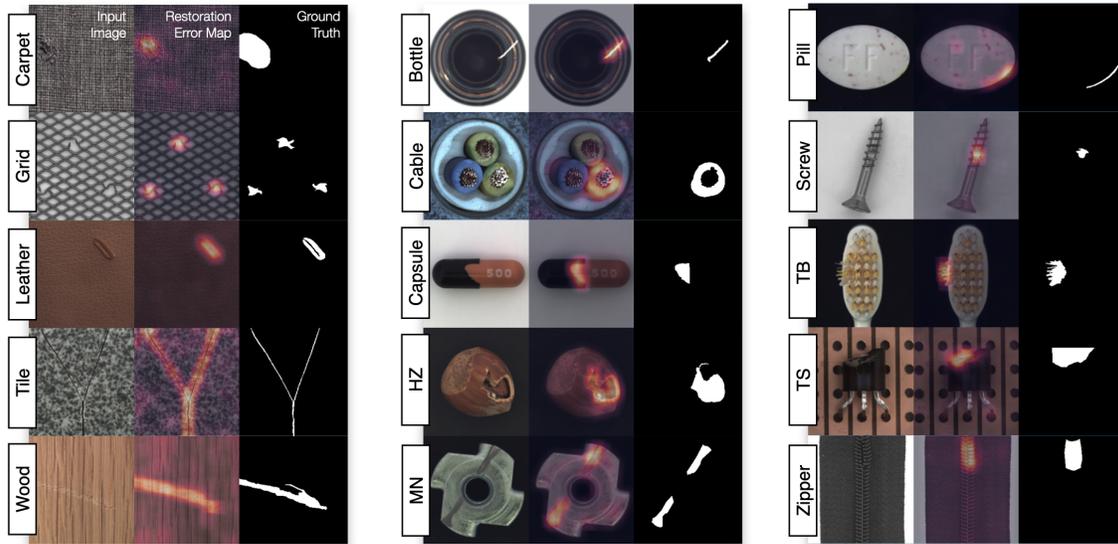
Figure 4. Examples of anomaly localization in MVTec AD using proposed method. The category names Hazelnut, Metal Nut, Toothbrush and Transistor are abbreviated as HN, MN, TB and TS respectively. From left to right: input image, restoration error map, ground truth. More examples of anomaly localization can be found in the supplementary material.

5.5. Results for Anomaly Localization

In the bottom part of Table 1, We compare the anomaly localization performance on MVTec AD between our method and others[7, 37, 50, 48, 17, 29]. In terms of pixel-level AUROC, our method reaches 97.4% and outperforms the previous top performer NSA [29].

In particular, the proposed method outperforms RIAD [50], the top-performing reconstruction-based method, by 3.2%p in terms of pixel-level AUROC. The anomaly localization examples of the different types of defects included in MVTec AD are shown in Figure 4. More anomaly localization examples can be found in the supplementary material.

The top part of Table 1 shows the comparison of our method with the state-of-the-art methods. PatchCore [25] and RD [8] use pre-trained models. DRÆM [49] uses an additional dataset for creating synthetic anomalies instead of the pre-trained model. Looking at the results, it seems that external knowledge helps to improve the localization performance. However, in general, since learned features of the pre-trained model are biased on the dataset used for pre-training, it is not free to use the model. In the unsupervised setting, we assume that there is no prior knowledge. Therefore, extra data to help improve performance is also unavailable. In this respect, the proposed approach showing competitive results without external knowledge is very meaningful.

We also evaluate the anomaly localization performance of the proposed method on BTAD and MPDD, and report the results in Table 2 and Table 3. Our method achieves pixel-level AUROC of 96.0% and 96.8% on BTAD and MPDD. The results demonstrated the effectiveness of the restoration-based approach for anomaly localization showing the best performance for product 01 and product 02 in BTAD and most categories in MPDD. In addition, the performance is higher than previous works using external knowledge, such as PatchCore [25], PaDiM [6], and SPADE [5]. Summarizing the results, the high anomaly localization performance of the proposed model is unconfined to a specific dataset. The result images of anomaly localization for MPDD and BTAD can be found in the supplementary material.

5.6. Results for Anomaly Detection

Comparing our method to state-of-the-art methods, such as PatchCore [25], DSR [51], and DRÆM [49], which get help from external knowledge is unfair. Therefore, we focus on the comparison with methods trained from scratch. Table 4 shows the anomaly detection results on MVTec AD. Our method achieves an image-level AUROC of 97.7%. This result is 0.5%p higher in image-level AUROC compared to NSA [29].

6. Discussion

Multiple Prediction. The proposed method requires the number of predictions N for multiple prediction to estimate the restoration error. While multiple prediction alleviates uncertainty caused by random perturbation, they increase computational cost. In Table 5, we report the anomaly de-

Method	Average	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule	HZ	MN	Pill	Screw	TB	TS	Zipper
						With external knowledge										
Metaformer* [46]	95.8	94.0	85.9	99.2	99.0	99.2	99.1	97.1	87.5	99.4	96.2	90.1	97.5	100	94.4	98.6
DRÆM [49]	98.0	97.0	99.9	100	99.6	99.1	99.2	91.8	98.5	100	98.7	98.9	93.9	100	93.1	100
DSR [51]	98.2	100	100	100	100	96.3	100	93.8	98.1	95.6	98.5	97.5	96.2	99.7	97.8	100
PatchCore [25]	99.1	98.7	98.2	100	98.7	99.2	100	99.5	98.1	100	100	96.6	98.1	100	100	99.4
						Without external knowledge										
TrustMAE* [37]	90.8	97.4	99.1	95.1	97.3	99.8	97.0	85.1	78.8	98.5	76.1	83.3	82.4	96.9	87.5	87.5
RIAD* [50]	91.7	84.2	99.6	100	98.7	93	99.9	81.9	88.4	83.3	88.5	83.8	84.5	100	90.9	98.1
Patch-SVDD [25]	92.1	92.9	94.6	90.9	97.8	96.5	98.6	90.3	76.7	92.0	94.0	86.1	81.3	100	91.5	97.9
CutPaste [17]	96.1	93.9	100	100	94.6	99.1	98.2	81.2	98.2	98.3	99.9	94.9	88.7	99.4	96.1	99.9
NSA [29]	97.2	95.6	99.9	99.9	100	97.5	97.7	94.5	95.2	94.7	98.7	99.2	90.2	100	95.1	99.8
Ours	97.7	91.7	100	99.9	99.8	96.1	100	94.2	97.2	98.6	96.6	96.1	98.6	98.1	98.7	99.9

Table 4. Performance comparison of anomaly detection in image-level AUROC(%) on the MVTec AD. For clarity, the category names Hazelnut, Metal Nut, Toothbrush and Transistor are abbreviated as HN, MN, TB and TS respectively. *: reconstruction-based method

	Image-level AUROC(%)	Pixel-level AUROC(%)
N = 1	97.5	97.3
N = 3	97.7	97.4
N = 5	97.7	97.4

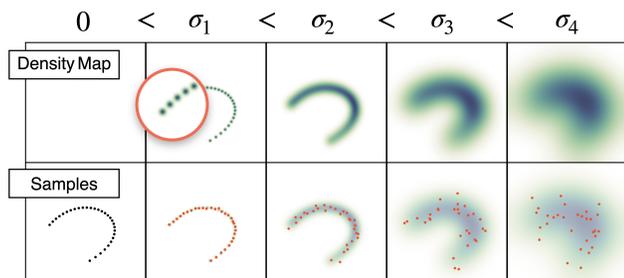Table 5. Results of the anomaly detection and localization performance with different N on MVTec AD.

Figure 5. Example showing a change of the perturbed data distribution depending on variance scale.

Figure 6. Radar chart showing the image-level detection performances depending on a variance scale parameter k. (Red line : optimized k for each class using log-scale grid search. Blue line : arbitrarily selected $k = 0.001$.)

tection and anomaly localization performance with different N on MVTec AD. The result shows that the detection performance is not significantly affected by N. Furthermore, the best performance can be reached by just using $N \geq 3$.

Effective Region. As shown in Figure 5, too small or large a variance scale prevents the distribution of perturbed data from reflecting the structural characteristics of the underlying manifold. From this result, we infer that there exists an effective region where the approximated score function works well depending on a variance scale because the approximation by denoising score-matching is affected by the variance scale. We compare the anomaly detection performance of applying the optimal k for each category and applying the empirically selected $k = 0.001$ (see Figure 6). In the case of applying the optimal k, the image-level
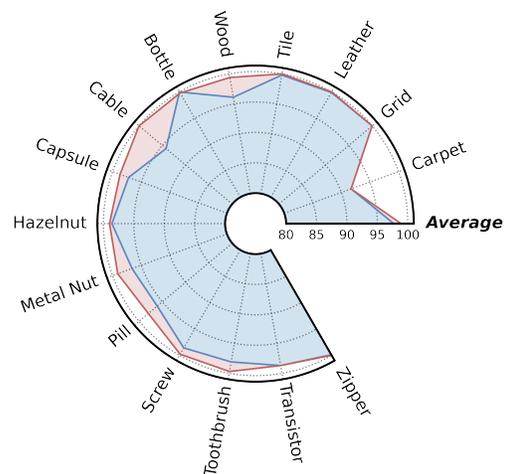
AUROC of some categories, such as bottle, wood, hazelnut, and metal nut, is improved. Also, the average image-level AUROC increases to 98.9%. Therefore, we will study to develop a method to find the effective region in future work.

7. Conclusion

In this paper, we proposed a novel unsupervised approach using a score-based model for anomaly detection. Our proposed method detects anomalies by identifying the samples that deviate from the normal manifold. In order to accomplish this, we leverage the score-based perturbation resilience for the sample. We defined score-based resilience to perturbation as the ability of the score to restore a perturbed sample to its original state instantly. Specifically, the variation of resilience depending on the sample position

was used as an indicator for anomaly detection. In addition, we derived this statement from the geometric perspective. The effectiveness of the proposed method for anomaly detection and localization was demonstrated, showing the outperforming results on the industrial anomaly detection benchmarks compared to the previous studies. Notably, considering that our proposed method does not rely on external knowledge such as pre-trained models or extra data, the proposed method has the potential to be widely scalable.

8. Acknowledgements

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018. 2

[2] Samet Akçay, Amir Atapour-Abarghouei, and T. Breckon. Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019. 2, 6

[3] Yoshua Bengio and Martin Monperrus. Non-local manifold tangent learning. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004. 3

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9584–9592, 2019. 5, 6

[5] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 3, 6, 7

[6] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021. 3, 6, 7

[7] David Dehaene, Oriel Frigo, Sébastien Combrexelle, and Pierre Eline. Iterative energy-based projection on a normal data manifold for anomaly localization. In *International Conference on Learning Representations*, 2020. 2, 6, 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. 6, 7

[9] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1705–1714, 2019. 2

[10] Matthäus Heer, Janis Postels, Xiaoran Chen, Ender Konukoglu, and Shadi Albarqouni. The ood blind spot of unsupervised anomaly detection. In *Medical Imaging with Deep Learning*, pages 286–300. PMLR, 2021. 2

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. 3, 4

[12] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. 6:695–709, dec 2005. 4

[13] Eric Jardim, Lucas A Thomaz, Eduardo AB da Silva, and Sergio L Netto. Domain-transformable sparse representation for anomaly detection in moving-camera videos. *IEEE Transactions on Image Processing*, 29:1329–1343, 2019. 2

[14] 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. 5, 6

[15] Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15425–15434, 2021. 2

[16] Jin-Hwa Kim, Do-Hyeong Kim, Saehoon Yi, and Taehoon Lee. Semi-orthogonal embedding for efficient unsupervised anomaly segmentation. *arXiv preprint arXiv:2105.14737*, 2021. 6

[17] 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. 3, 6, 7, 8

[18] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2152–2161, 2019. 2

[19] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J Radke, and Octavia Camps. Towards visually explaining variational autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8642–8651, 2020. 2

[20] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018. 6

[21] Hui Lv, Chuanwei Zhou, Zhen Cui, Chunyan Xu, Yong Li, and Jian Yang. Localizing anomalies from weakly-labeled videos. *IEEE transactions on image processing*, 30:4505–4515, 2021. 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. 5, 6

[23] Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022. 2

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 5

[25] 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 (CVPR)*, pages 14318–14328, June 2022. 3, 5, 6, 7, 8

[26] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. 3

[27] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018. 2

[28] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 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 Nature Switzerland Cham, 2022. 3, 5, 6, 7, 8

[30] Philipp Seeböck, José Ignacio Orlando, Thomas Schlegl, Sebastian M Waldstein, Hrvoje Bogunović, Sophie Klimscha, Georg Langs, and Ursula Schmidt-Erfurth. Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal oct. *IEEE transactions on medical imaging*, 39(1):87–98, 2019. 2

[31] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 3

[32] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. 3

[33] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *CoRR*, abs/1907.05600, 2019. 4

[34] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *CoRR*, abs/2011.13456, 2020. 3, 4

[35] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 5

[36] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6479–6488, 2018. 2

[37] Daniel Stanley Tan, Yi-Chun Chen, Trista Pei-Chun Chen, and Wei-Chao Chen. Trustmae: A noise-resilient defect classification framework using memory-augmented autoencoders with trust regions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 276–285, 2021. 2, 6, 7, 8

[38] 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, Jun 2020. 6

[39] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. 3

[40] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *European Conference on Computer Vision*, pages 485–503. Springer, 2020. 2

[41] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, jul 2011. 4

[42] Pascal Vincent and Yoshua Bengio. Manifold parzen windows. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2002. 3

[43] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. ICML '08, page 1096–1103, New York, NY, USA, 2008. Association for Computing Machinery. 3

[44] LeiChen Wang, Simon Giebenhain, Carsten Anklam, and Bastian Goldluecke. Radar ghost target detection via multimodal transformers. *IEEE Robotics and Automation Letters*, 6(4):7758–7765, 2021. 2

[45] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C. Cattin. Diffusion models for medical anomaly detection. In Linwei Wang, Qi Dou, P. Thomas Fletcher, Stefanie Speidel, and Shuo Li, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 35–45, Cham, 2022. Springer Nature Switzerland. 2

[46] Jhih-Ciang Wu, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Learning unsupervised metaformer for anomaly

detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4369–4378, October 2021. 2, 6, 8

[47] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 649–655, 2022. 2

[48] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020. 2, 6, 7

[49] 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. 3, 6, 7, 8

[50] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021. 2, 6, 7, 8

[51] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *European Conference on Computer Vision*, pages 539–554. Springer, 2022. 6, 7, 8

[52] Ye Zheng, Xiang Wang, Rui Deng, Tianpeng Bao, Rui Zhao, and Liwei Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2022. 6

[53] David Zimmerer, Fabian Isensee, Jens Petersen, Simon Kohl, and Klaus Maier-Hein. Unsupervised anomaly localization using variational auto-encoders. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 289–297. Springer, 2019. 2