Frequency-dependent Image Reconstruction Error for Micro Defect Detection

Yuhei Nomura Y-NOMURA@WAKAYAMA-KG.JP
Industrial Technology Center of Wakayama Prefecture, Japan
Graduate School of Systems Engineering, Wakayama University, Japan

Hirotaka Hachiya HHACHIYA@WAKAYAMA-U.AC.JP
Graduate School of Systems Engineering, Wakayama University, Japan

Editors: Berrin Yanıkoğlu and Wray Buntine

Abstract

Micro defects, such as casting pores in industrial products, have been detected by human visual inspection using X-ray CT images and image processing tools. Automatic detection of micro defects is challenging for anomaly detection methods using image reconstruction errors and nearest neighbor distances because these metrics are dominated by low-frequency information and are insensitive to minor defects. Although recent methods achieve high anomaly detection performances, their detection abilities are insufficient for micro defects. To overcome these problems, we propose to extend a state-of-the-art anomaly detection method by introducing frequency-dependent losses to capture reconstruction errors appearing around micro defects and frequency-dependent data augmentation to improve the sensitivity against the errors. We demonstrate the effectiveness of the proposed method through experiments with MVTec AD dataset especially on the detection of micro defects.
Keywords: industrial inspection; anomaly detection; deep network

1. Introduction

Internal defects such as casting pores may be formed during the manufacturing processes of industrial products, such as castings and plastic. These defects can be observed by non-destructive inspection using industrial X-ray computed tomography (CT), but these defects are often tiny, and those CT images tend to be micro and unsharp. In the case of high-mix low-volume production, inspection targets have various shapes and are composed of various materials. Also, adjusting radiographic conditions each time to acquire clear images is costly. Therefore, detecting micro defects in industrial CT images often requires human decision-making, and thus automatic detection is highly required.

An anomaly or outlier detection method is generally used for defect detection because of the low availability of anomaly data. Several anomaly detection methods based on deep learning have recently been proposed for visual inspection. These methods can be classified into two categories, feature-extraction-based and reconstruction-based. Feature-extraction-based methods, e.g., SPADE (Cohen and Hoshen, 2020) and PatchCore (Roth et al., 2022), extract pyramid features from pre-trained image classification model and

detect defects based on nearest neighbor distances from normal deep features. However, detecting micro defects which vary in a small region, e.g., at the pixel level, is difficult because pyramid features are obtained by multi-scale abstraction through convolutional operations. On the other hand, reconstruction-based methods, e.g., DRÆM (Zavrtanik et al., 2021) and OCR-GAN (Liang et al., 2022), use Autoencoder (AE) or Generative Adversarial Networks (GANs) which are trained to reconstruct only normal images and detect defects based on reconstruction errors. In this way, discrimination at the pixel level is possible in principle because the resolution of the reconstructed image is not degraded. However, reconstructed images mostly lose high-frequency information due to repeated convolution operations. Although the quality of reconstructed images has been improved by advanced loss functions such as Structural Similarity (SSIM) loss (Bergmann et al., 2019), it is not enough to detect micro defects.

The above methods have achieved extremely high performances in both anomaly detection and localization, therefore detection of micro defects seems not to be a problem in existing methods. However, the dataset used for their evaluation, MVTec AD (Bergmann et al., 2021), only contains defects relatively larger than real CT images, and the detection ability of micro defects has not been discussed in recent works. Meanwhile, transforming images into the frequency domains with Fourier transform or wavelet transform would be adequate for defect detection (Czimmermann et al., 2020). Therefore, assuming that the changes in micro defects can be captured in the high-frequency domain, we propose introducing frequency-dependent reconstruction error and data augmentation to detect micro defects.

The main contributions of this paper are summarized as follows:

1. We propose a new anomaly detection framework for detecting micro anomalies based on frequency-dependent reconstruction error. This framework enables dynamic integration of reconstruction errors in optimal frequency domains.

2. We propose a new data augmentation method adding noises to frequency-separated images. This method assists the training in detecting high-frequency anomalies.

3. We conduct extensive comparative experiments on anomaly detection with MVTec AD dataset, demonstrating that the proposed method outperforms existing methods.

After this introductory section, the rest of this paper is organized as follows. Section 2 describes the formulation of problems for reconstruction-based image anomaly detection and reviews related works including reconstruction-based image anomaly detection methods. Section 3 details the proposed method. Section 4 describes the experimental evaluation and discussion, and the conclusion is presented in Section 5.

2. Related works

In this section, we describe the formulation of problems for reconstruction-based image anomaly detection and review related works.

2.1. Problem formulation

Let $I \in \mathbb{R}^{C\times H\times W}$ and $M \in \mathbb{R}^{C\times H\times W}$ be an original image and its anomaly mask, where C denotes the number of channels of the image, H and W are the image height and width. In the case of reconstruction-based methods, the reconstructed image $\widehat{I} \in \mathbb{R}^{C\times H\times W}$ is estimated from I with a reconstruction function $\mathrm{rec}(\cdot)$, which is trained to restore only normal images correctly. Finally, the estimated anomaly mask $\widehat{M} \in \mathbb{R}^{H\times W}$ is obtained based on a pair of images I and $\widehat{I}$, e.g., using reconstruction error between I and $\widehat{I}$, by a scoring function $\mathrm{score}(\cdot)$ as follows:

$$\widehat{M} = \mathrm{score}(I, \widehat{I}) = \mathrm{score}(I, \mathrm{rec}(I)). \tag{1}$$

To generate high scores for micro defects in $\widehat{M}$, the reconstruction function $\mathrm{rec}(\cdot)$ is required to generate a fine-grained reconstructed image $\widehat{I}$, containing accurate high-frequency information. In addition, the scoring function $\mathrm{score}(\cdot)$ is required to detect the difference between I and $\widehat{I}$ in the high-frequency domain. However, in general, Images I comprises a large proportion of low-frequency information; therefore, building such $\mathrm{rec}(\cdot)$ and $\mathrm{score}(\cdot)$ is difficult.

2.2. Discriminative reconstruction error

So far, various hand-made scoring functions $\mathrm{score}(\cdot)$ combining multiple metrics, e.g., L1 and L2 errors, have been used for measuring reconstruction error, but a standard function does not exist. In the work (Zavrtanik et al., 2021), a new framework, called DRÆM, has been proposed to train scoring function in a supervised manner. More specifically, training data consist of pseudo-anomaly images generated by the data augmentation method, randomly generating masks based on Perlin noise and synthesizing the external images. Thus, it enables the scoring function to detect anomalies discriminatively and achieves the highest performance in an industrial inspection dataset, MVTec AD (Bergmann et al., 2021). However, reconstructed images $\widehat{I}$ can lose high-frequency information due to the property of convolutional operations. In addition, since pseudo-anomaly images are assumed to be dominated by low-frequency information, the discriminative network can not learn high-frequency features. For these reasons, discriminating micro defects is still challenging even with trained scoring functions $\mathrm{score}(\cdot)$.

2.3. Reconstruction with frequency separation

The work (Liang et al., 2022) proposed the framework, called OCR-GAN, which separates an input image into multiple frequency images using a Laplacian pyramid, and reconstructs them individually by each network through sharing features using channel attention. This method aims to improve the quality of high-frequency information. However, detecting high-frequency anomalies including micro defects would be difficult because the reconstruction error is calculated for the single image in which all frequency ones are aggregated and dominated by low-frequency information. In addition, this method is available only for image-level anomaly detection and does not support the localization of anomalies, i.e., a pixel-level anomaly.

2.4. SSIM loss

Image reconstruction by AE with L1/L2 loss function tends to lose high-frequency information and reconstructed images become blurred. For the anomaly detection task, SSIM loss (Bergmann et al., 2019) is typically used to improve the quality of reconstructed images. SSIM index between two patches X and Y is calculated using these means μ_* and standard deviations σ_* where $* \in \{X, Y\}$ as follows:

$$\mathrm{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{({\mu_X}^2 + {\mu_Y}^2 + C_1)({\sigma_X}^2 + {\sigma_Y}^2 + C_2)}, \tag{2}$$

where C_1 and C_2 are parameters to avoid the zero division. Typically, let the user-defined parameter and the dynamic range of the pixel values be k_* and L_*, they are denoted as $C_1 = (k_1 L_1)^2$ and $C_2 = (k_2 L_2)^2$. Then, SSIM loss is obtained by the average of SSIM indices of corresponding patches of original I and reconstructed $\widehat{I}$ images as follows:

$$\mathcal{L}_{\mathrm{SSIM}}(I, \widehat{I}) = \frac{1}{P}\sum_{p=1}^{P} 1 - \mathrm{SSIM}(I_p, \widehat{I}_p), \tag{3}$$

where I_p and $\widehat{I}_p$ represent p-image patches split by sliding windows with a stride of 1, and P is the number of patches in each image.

However, because the SSIM index is computed based on abstraction, i.e., the mean and variance of the values in a patch, it tends to lose information at high frequencies and would not be enough to reveal micro defects.

2.5. Spatial Frequency Loss

For fine-grained AE and resulting feature-extraction, Ichimura (2018) proposed Spatial Frequency Loss (SFL) which allows setting greater weights to the reconstruction errors in the higher frequency components of an image. SFL is calculated with mean squared error (MSE) between input image I and reconstructed image $\widehat{I}$ in multiple-frequency levels as follows:

$$\begin{aligned} \mathcal{L}_{\mathrm{SFL}}(I, \widehat{I}) &= \sum_{s=0}^{S-1} w_s E_{\mathrm{SFL}}(I, \widehat{I}, \boldsymbol{\alpha}_s), \\ E_{\mathrm{SFL}}(I, \widehat{I}, \boldsymbol{\alpha}_s) &= \frac{1}{CWH}\left\|\mathrm{freq}(I, \boldsymbol{\alpha}_s) - \mathrm{freq}(\widehat{I}, \boldsymbol{\alpha}_s)\right\|_F^2, \end{aligned} \tag{4}$$

where $\mathrm{freq}(I, \boldsymbol{\alpha}_s) \in \mathbb{R}^{C\times W\times H}$ is a function to generate an image containing only components at a specific frequency represented by a parameter $\boldsymbol{\alpha}_s$, e.g., frequency bands, from image I. S is the number of frequency levels, $w_s \in \mathbb{R}$ is the weight of s-th frequency band, and $\|\cdot\|_F$ is Frobenius norm. With a larger w_s value for a higher frequency level, SFL can enable AE to reconstruct high-frequency information accurately.

3. Proposed method

To improve anomaly detection for micro defects, fine-grained reconstruction and frequency-dependent measures of reconstruction error are required. To realize them, we propose a

framework called FIRE-AD (Frequency-dependent Image Reconstruction Error for Anomaly Detection), which separates original and reconstructed images into multiple frequency images, measures reconstruction errors in each frequency band, and aggregates them dynamically to generate an anomaly score map. In addition, to enhance discriminative detection for micro defects, we propose frequency-dependent data augmentation in which pseudo-anomalies are added to frequency-separated images directly.

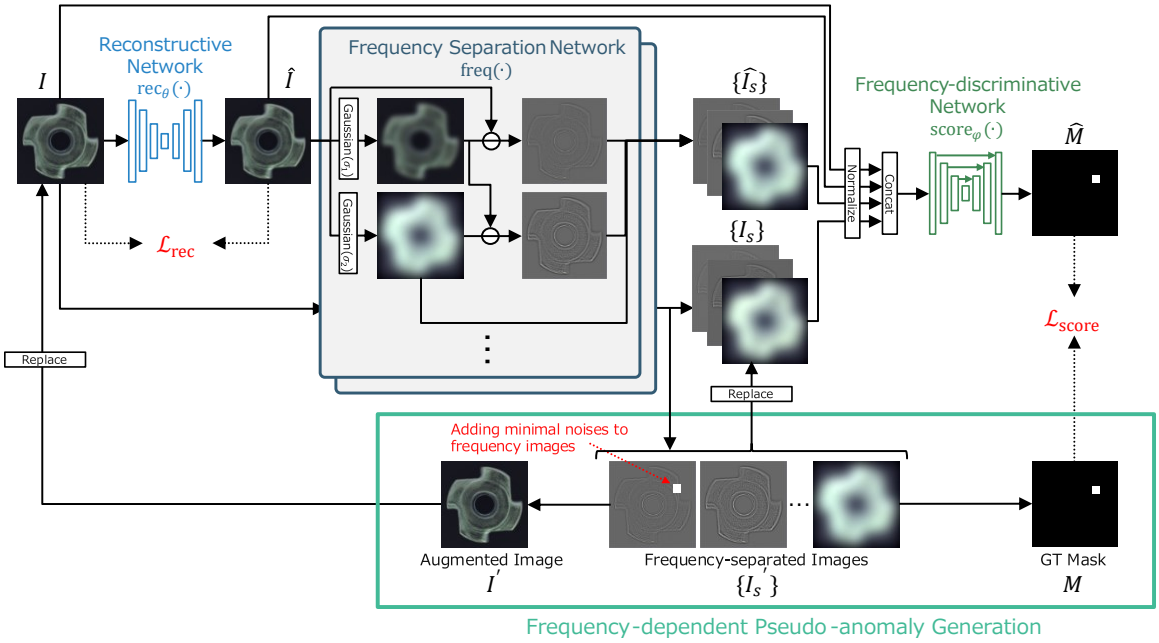


Figure 1: Architecture overview of FIRE-AD

3.1. FIRE-AD

The architecture overview of the proposed FIRE-AD is shown in Fig. 1. FIRE-AD comprises three main components: reconstructive network $\mathrm{rec}_{\boldsymbol{\theta}}(\cdot)$, frequency separation network $\mathrm{freq}(\cdot)$, and frequency-dependent discriminative network $\mathrm{score}_{\boldsymbol{\phi}}(\cdot)$.

The reconstructive network $\mathrm{rec}_{\boldsymbol{\theta}}(\cdot)$ is a generative network, i.e., AE, and its reconstruction ability for high-frequency information is enhanced by training with a frequency weighting loss function, i.e., SFL. The frequency separation network $\mathrm{freq}(\cdot)$ takes the role of image separation into frequency domains. This makes input image I and reconstructed image $\widehat{I}$ separated into multiple frequency images as follows:

$$\begin{aligned} \mathcal{I} &\equiv \{\mathrm{freq}(I, \boldsymbol{\alpha}_s) | s = 0, 1, 2, \ldots, S-1\}, \\ \widehat{\mathcal{I}} &\equiv \{\mathrm{freq}(\widehat{I}, \boldsymbol{\alpha}_s) | s = 0, 1, 2, \ldots, S-1\}. \end{aligned} \tag{5}$$

The frequency-dependent discriminative network generates anomaly score map $\widehat{M}$ based on I, $\widehat{I}$, $I_s \in \mathcal{I}$, and $\widehat{I}_s \in \widehat{\mathcal{I}}$ as follows:

$$\widehat{I} = \mathrm{rec}_{\boldsymbol{\theta}}(I), \quad \widehat{M} = \mathrm{score}_{\boldsymbol{\phi}}(I, \widehat{I}, \mathcal{I}, \widehat{\mathcal{I}}), \tag{6}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are parameters of networks. Finally, the anomaly score of the image $\widehat{A}$ is calculated from $\widehat{M}$ with the method proposed in Zavrtanik et al. (2021), as follow:

$$\widehat{A} = \max(\widehat{M} * f_{\text{mean}}), \tag{7}$$

where f_{mean} is a mean filter and its size is set to 21×21.
The reconstructive network $\text{rec}_{\boldsymbol{\theta}}(\cdot)$ is trained by minimizing L2 and SFL losses as follows:

$$\mathcal{L}_{\text{rec}}(I, \widehat{I}) = \mathcal{L}_2(I, \widehat{I}) + \lambda_{\text{SFL}}\mathcal{L}_{\text{SFL}}(I, \widehat{I}). \tag{8}$$

The discriminative network $\text{score}_{\boldsymbol{\phi}}(\cdot)$ is trained with focal loss (Lin et al., 2017) due to imbalanced classes and the total loss for training is as follows:

$$\begin{aligned} \mathcal{L}_{\text{score}}(M, \widehat{M}) &\equiv \lambda_{\text{focal}}\mathcal{L}_{\text{focal}}(M, \widehat{M}), \\ \mathcal{L}_{\text{total}} &\equiv \mathcal{L}_{\text{rec}}(I, \widehat{I}) + \mathcal{L}_{\text{score}}(M, \widehat{M}), \end{aligned} \tag{9}$$

where λ_* is a weight for each loss.

3.2. Frequency separation network

The frequency separation network $\text{freq}(\cdot)$ connects the reconstructive and discriminative networks; therefore, the gradients must propagate through the frequency separation network. For frequency image separation, Fourier or wavelet transform is generally used; however, it is not straightforward to pass gradients through them. To overcome this problem, we propose to use DoG (Different of Gaussian) filters, which consist of 2D-convolution operations allowing gradient propagation. Specifically, a DoG filter calculates the difference between two images filtered by Gaussian filters with different standard deviations σ. Since a DoG filter works equivalent to a band-pass filter, frequency-separated images $I_s \in \mathbb{R}^{C \times H \times W}$ are obtained by repeated DoG filters while changing σ step by step as follows:

$$I_s = \text{freq}(I, \boldsymbol{\alpha}_s) = I * K(\sigma_s) - I * K(\sigma_{s+1}), \tag{10}$$

$$K(\sigma) = \begin{cases} \mathbf{1} & \text{if } \sigma = 0, \\ \mathbf{0} & \text{if } \sigma = \infty, \\ G(\sigma) & \text{otherwise}, \end{cases} \tag{11}$$

where K and $*$ indicate a kernel matrix and convolution operation, respectively. $\mathbf{0} \in \mathbb{R}^{1\times1}$ and $\mathbf{1} \in \mathbb{R}^{1\times1}$ are kernel matrices of 0s and 1s, respectively. $\boldsymbol{\alpha}_s = \{\sigma_s, \sigma_{s+1}\}$ is a pair of frequency band parameters σ and $G(\sigma) \in \mathbb{R}^{k\times k}$ is a Gaussian kernel with standard deviation σ. In addition, to increase the bandwidth of lower frequency, the size k of Gaussian kernel is varied according to σ as follows:

$$k' = \lfloor 8\sigma \rfloor, \quad k = k' + 1 - (k' \bmod 2). \tag{12}$$

We note that these operations are applied per channel; because the color information of RGB is important to detect colored defects.

3.3. Frequency-dependent pseudo-anomaly generation

As mentioned in Sec. 2.2, the pseudo-anomaly generation proposed in DRÆM does not consider producing high-frequency anomalies; therefore, the discriminative network cannot extract useful features, e.g., from high-frequency appearance, to detect micro defects.

Therefore, we propose a novel argumentation method directly enhancing high-frequency information by adding micro noises to frequency images. The diagram of the augmentation process is shown in Fig. 2. At first, we prepare normal frequency images $\{I_s\}$ from a normal image I. For each I_s, a binary anomaly mask $M_s \in \{0,1\}^{H\times W}$ is generated by adding micro noises, i.e., the values of 1, at random positions. Then, the augmented anomaly frequency image I'_s, the augmented anomaly image I' and the ground truth mask M are generated as follow:

$$I'_s = I_s \odot M_s, \quad I' = \sum_{s=0}^{S-1} I'_s, \quad M[h,w] = \mathbb{1}\Big(\sum_{s=0}^{S-1} M_s[h,w] \geq 1\Big), \tag{13}$$

where $*[h,w]$ indicates the value of pixel (h,w) in image $*$ and $\mathbb{1}(x)$ is an indicator function, which returns 1 when the condition x is true and 0 otherwise.

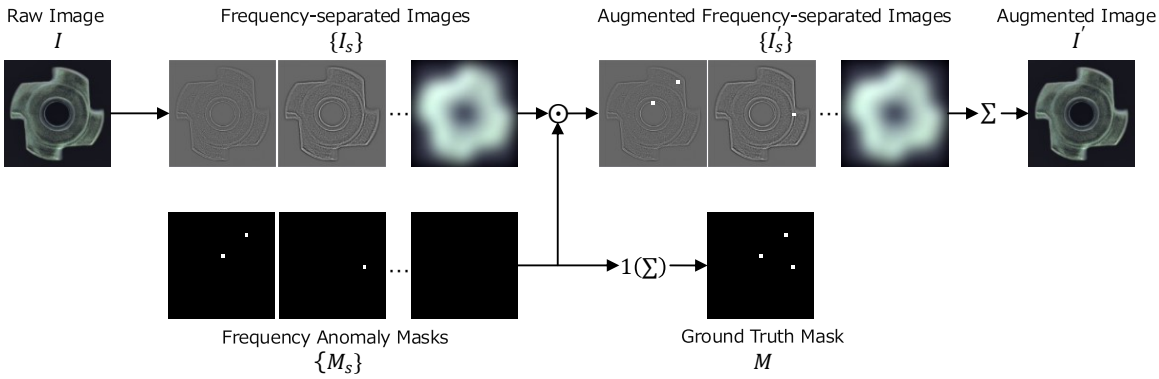


Figure 2: Schematic diagram of frequency-dependent pseudo-anomaly generation

4. Experimental evaluation

In this section, we show the effectiveness of the proposed methods through experiments with MVTec AD (Bergmann et al., 2021) dataset. Since MVTec AD contains mainly large anomalies and is not suitable for evaluating the performance of detecting micro defects, we also use synthetic micro defect dataset based on MVTec AD.

4.1. Datasets

Let $\mathcal{D}^{\mathrm{tr}}$ and $\mathcal{D}^{\mathrm{te}}$ be pairs of image I and its ground-truth anomaly map M as follows:

$$\mathcal{D}^{\mathrm{tr}} \equiv \Big\{\big(I_i, M_i\big)\Big\}_{i=1}^{N^{\mathrm{tr}}}, \quad \mathcal{D}^{\mathrm{te}} \equiv \Big\{\big(I_i, M_i\big)\Big\}_{i=1}^{N^{\mathrm{te}}}, \tag{14}$$

where N^{tr} and N^{te} are the training and test data numbers, respectively.

MVTec AD (Bergmann et al., 2021) A standard benchmark dataset for industrial image anomaly detection. There are 15 different objects, i.e., leather and metal nut, each with a pair of training $\mathcal{D}^{\text{tr}}$ and test $\mathcal{D}^{\text{te}}$ data. $\mathcal{D}^{\text{tr}}$ consists of only normal data with N^{tr} from 60 to 320 depending on objects, and $\mathcal{D}^{\text{te}}$ consists of 12 to 60 normal and 30 to 141 abnormal data depending on objects. We note that the size of all images and masks in training and test data is resized to $(H, W) = (256, 256)$ to make the conditions the same as existing methods (Zavrtanik et al., 2021; Liang et al., 2022). We used this dataset to compare the proposed and existing methods to evaluate general anomaly detection performances.

Synthetic micro defect dataset We generated synthetic micro defects dataset to evaluate the anomaly detection performance for micro defects. This dataset is only for testing; normal data is the same as original MVTec AD and anomaly one is replaced to the synthetic anomaly data. More specifically, we randomly selected 50 normal data from $\mathcal{D}^{\text{te}}$ of each object while allowing for duplication, resized to $(H, W) = (256, 256)$, and added gray-scale dot noises at random positions—the pixel value of noises is uniformly selected in the range of 50 to 200 to make detecting dots difficult. Examples of synthetic micro defects are depicted in Fig. 3, showing that anomaly dots are distributed over the entire region of an image, including the object and its background, and detecting anomalies located over the edges of the object is difficult.

4.2. Experimental settings

We compared the performance of state-of-the-art anomaly detection methods, PatchCore, DRÆM, OCR-GAN, and the following two proposed methods:

- PatchCore (Roth et al., 2022)—we used the official implementation and settings of ensemble model. The size of images for both training and testing is resized to $(H, W) = (256, 256)$ and not center-cropped, to make the conditions the same as other methods.

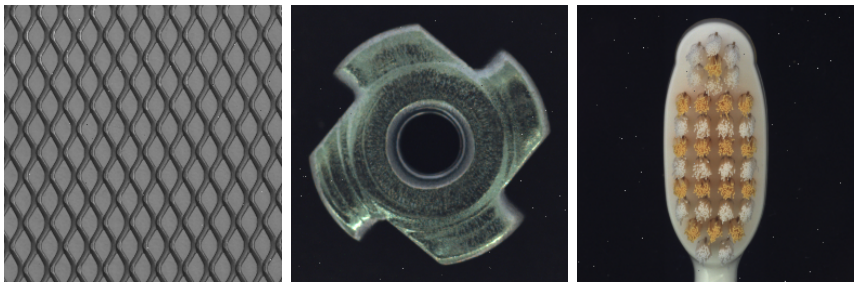
- DRÆM (in Sec. 2.2)—we used the same implementation as our FIRE-AD for retraining DRÆM with the number of frequency levels $S = 0$, meaning that frequency-seperation is not performed. Instead of Eq. 8, the reconstructive network was trained with SSIM loss (in Sec. 2.4) as follows:

$$\mathcal{L}_{\text{rec}}(I, \widehat{I}) = \mathcal{L}_2(I, \widehat{I}) + \lambda_{\text{SSIM}}\mathcal{L}_{\text{SSIM}}(I, \widehat{I}), \tag{15}$$

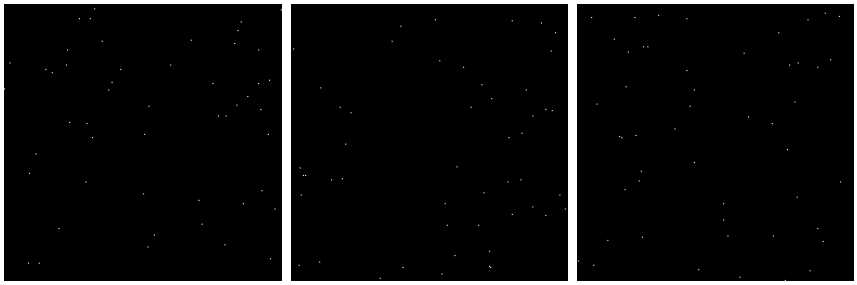
  where λ_{SSIM} is the weight for SSIM loss and set to 1—parameters of SSIM loss (see Eq. 2) are set as $k_1 = 0.01$ and $k_2 = 0.03$. In addition, the weight of focal loss λ_{focal} (in Eq. 9) was set to 1.

- OCR-GAN (in Sec. 2.3)—we used results provided by (Liang et al., 2022).

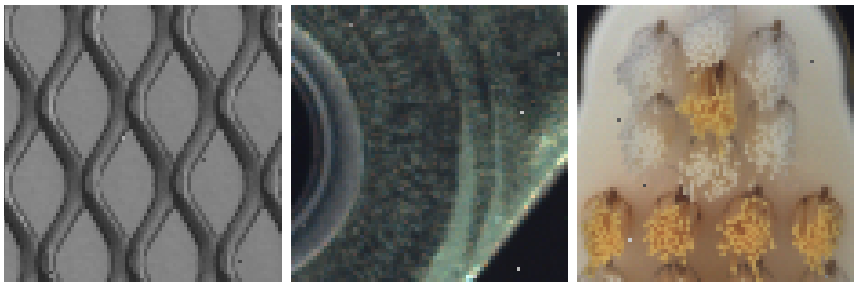
- FIRE-AD—proposed method (in Sec. 3.1). We set $S = 2$, $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \sigma_2) = (0, 0.6, \infty)$, and $\boldsymbol{w} = (w_0, w_1) = (1000, 10)$, for the parameters of the frequency separation network and SFL. The weights of SFL λ_{SFL} and focal loss λ_{focal} were both set



(a) Examples of synthesized anomaly images I



(b) Examples of corresponding ground truth masks M



(c) Enlarged view of (a)

Figure 3: Examples of synthetic micro defects with 1-px gray-scale dot noises.

to 1. The images I, $\widehat{I}$, $\mathcal{I}$ and $\widehat{\mathcal{I}}$ are normalized with Z-score normalization before input to the dicriminative network. To confirm the effectiveness of reconstruction error with high-frequency information, we used only high-frequency images, as $\mathcal{I} = \{I_0\}$ and $\widehat{\mathcal{I}} = \{\widehat{I}_0\}$ (in Eq. 5).

- FIRE-AD$_{\text{ag}}$—proposed method trained with frequency-dependent pseudo-anomaly generation (in Sec. 3.3). The augment target in frequency images is I_0, which is the highest frequency image, and the size and amount of added noises are set to 1-px and 100 for each image. The pixel value of noises is the max value of target I_0. The other settings are the same as FIRE-AD.

For training and testing of all models except PatchCore and OCR-GAN, we set the number of training epochs to 300, the batch size to 4, and the initial learning rate to 0.0001—the rate was multiplied by 0.2 at each of 240th and 270th epochs for the stability of training.

4.3. Evaluation metrics

Image-level Area Under the Receiver Operating Characteristic Curve (AUROC) and pixel-level AUROC are generally used to evaluate the performance of anomaly detection and localization, respectively (Cohen and Hoshen, 2020; Roth et al., 2022; Zavrtanik et al., 2021; Liang et al., 2022). However, pixel-level AUROC is insufficient for a fair evaluation of the localization performance because the number of anomalous pixels is much less than normal ones. In addition, pixel-level AUROC does not consider the size of anomaly regions; therefore, the score is dominated by large anomalies mainly composed of low-frequency information. For these reasons, the performance of anomaly localization is also evaluated with pixel-level Average Precision (AP) and Area Under Per-Region-Overlap (AUPRO) (Bergmann et al., 2021). PRO which enables to handle of different-sized anomalies as the same size by size normalization, is calculated as follows:

$$\text{PRO} = \frac{1}{N^{\text{reg}}} \sum_{i=1} \sum_{k} \frac{|P_i \cap C_{i,k}|}{C_{i,k}}, \tag{16}$$

where N^{reg} is the number of ground truth regions in the entire dataset, $C_{i,k}$ is the ground truth of positive k-th region of image I_i, and P_i is the predicted region of I_i. Finally, in the same manner as AUROC, AUPRO is obtained by calculating the area under PRO curve. Typically, AUPRO is calculated with the area in which False Positive Rate (FPR) is under 0.3 and normalized to [0,1].

4.4. Evaluation Results

The quantitative comparison of the image-level detection and pixel-level localization performances on the original MVTec AD are shown in Table 1 and Table 2, respectively. In addition, the qualitative comparisons are shown in Fig. 4. As mentioned in Sec. 2.3, OCR-GAN does not support anomaly localization; therefore, only the image-level detection performance is compared. Table 1 and Table 2 show that the performances of the proposed

Table 1: Anomaly detection performance (image-level AUROC%) on the MVTec AD dataset. The best score among all methods for each object is indicated in bold. We note that the performances of PatchCore and DRÆM are re-trained and re-evaluated on our experimental settings.

Class	PatchCore	DRÆM	OCR-GAN	FIRE-AD	FIRE-AD$_{ag}$
bottle	**100**	99.0	99.6	99.4	98.6
cable	**99.8**	94.0	99.1	91.9	93.0
capsule	**99.0**	93.7	96.2	88.4	91.3
carpet	98.7	97.8	**99.4**	96.3	90.2
grid	99.4	**100**	99.6	99.9	**100**
hazelnut	**100**	99.9	98.5	99.3	**100**
leather	**100**	**100**	97.1	**100**	**100**
metal nut	**100**	**100**	99.5	99.9	99.5
pill	97.8	98.1	**98.3**	83.3	84.2
screw	98.6	83.2	**100**	70.1	79.7
tile	**100**	**100**	95.5	**100**	99.9
toothbrush	92.7	**100**	98.7	**100**	**100**
transistor	**99.8**	92.8	98.3	89.5	87.3
wood	98.9	**100**	95.7	**100**	99.5
zipper	98.7	**100**	99.0	**100**	**100**
avg.	**98.9**	97.2	98.3	94.5	94.9

methods are well comparable with the ones of existing methods, PatchCore, DRÆM and OCR-GAN although the proposed methods are tuned to detect micro defects.

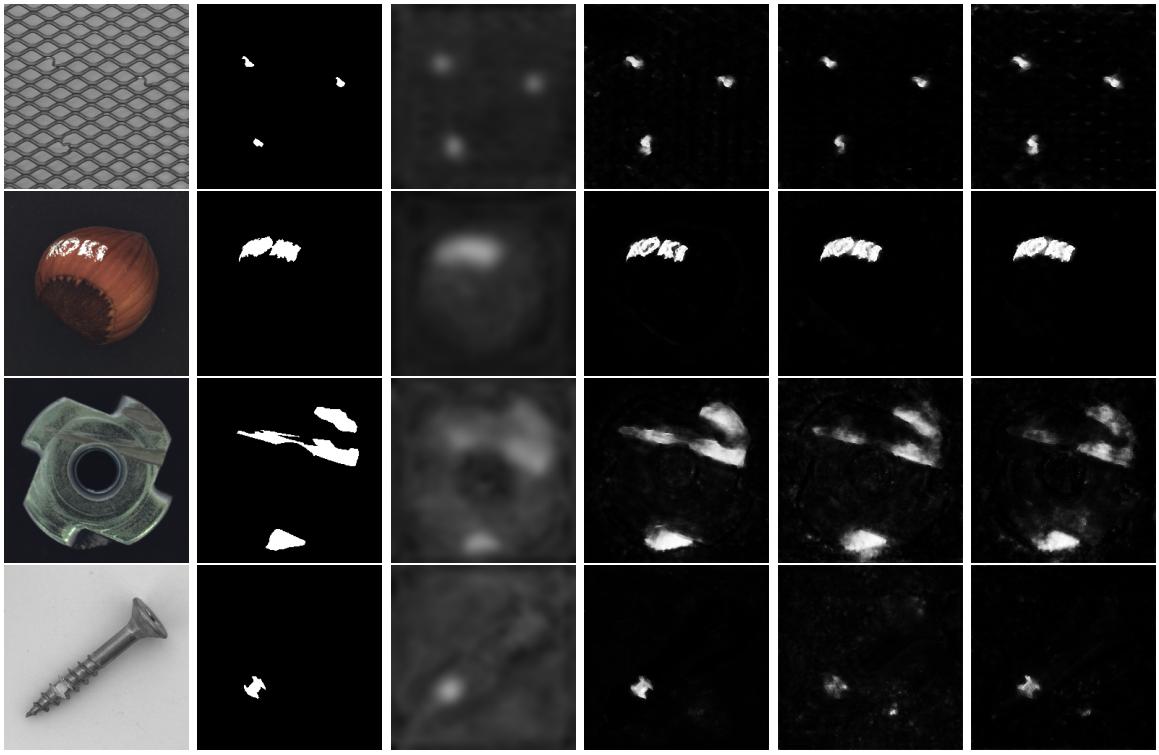
The quantitative comparison on the synthetic micro defect dataset are shown in Table 3 and Table 4, respectively. We note that pixel-level localization performances are not evaluated with AUPRO because all anomalies have the same size and AUPRO is equal to AUROC. The qualitative comparisons are also shown in Fig. 5. Since the synthetic micro defect dataset is highly imbalanced in pixel-level, the pixel-level AP score is of high importance. From the quantitative comparisons, PatchCore has the highest performance in detection but the lowest in localization. Especially, the pixel-level AP score is almost under 1%, indicating that PatchCore is not able to detect micro defects. Compared to DRÆM, our proposed FIRE-AD performs better in both the detection and localization performances on the micro defect dataset. These results suggest that measuring frequency-dependent reconstruction error is effective to detect micro defects. While the detection performance of proposed method with frequency-dependent augmentation FIRE-AD$_{ag}$ is lower than the others, the localization performance is almost the best. The high pixel-level AP score indicates that the proposed method tends to detect micro defects with high accuracy. Frequency-dependent augmentation may assist learning high-frequency features to improve discriminative ability, but also make sensitive to native noises. This can be observed in the qualitative comparisons Fig.5e and 5f; they are over responding to native patterns. It is considered important to suppress over-fitting to native noises.

Table 2: Anomaly localization performance (pixel-level AUROC%/AP%/AUPRO%) on the MVTec AD dataset. The best score among all methods for each object is indicated in bold. We note that the performances of PatchCore and DRÆM are re-trained and re-evaluated on our experimental settings.

Class	PatchCore	DRÆM	FIRE-AD	FIRE-AD$_{ag}$
bottle	**98.8**/76.4/**95.7**	97.9/**84.4**/94.8	96.0/76.4/90.5	94.6/76.7/90.2
cable	**98.8**/66.8/**95.1**	95.3/**68.5**/82.8	92.9/49.6/74.5	90.0/46.4/64.7
capsule	**99.2**/**45.5**/**95.9**	81.6/42.0/80.5	87.8/23.7/72.8	90.3/31.6/83.7
carpet	**99.0**/60.0/**94.6**	94.6/44.7/88.5	98.0/**77.2**/94.2	97.1/66.4/90.6
grid	98.7/26.5/94.7	**99.6**/**71.7**/**98.9**	99.2/56.3/97.9	**99.6**/64.7/98.3
hazelnut	**99.0**/53.8/96.1	98.6/**76.6**/**97.4**	98.7/75.9/96.0	98.7/75.8/96.8
leather	99.3/42.6/98.0	98.0/68.5/97.0	99.3/**68.7**/98.5	**99.4**/67.6/**98.8**
metal nut	98.9/89.2/95.9	**99.2**/**94.4**/**96.2**	98.5/90.6/92.1	98.6/90.4/91.2
pill	**98.4**/**80.4**/**94.9**	97.6/51.6/92.0	95.8/73.4/74.5	92.6/65.6/63.1
screw	**99.5**/35.9/**96.9**	97.2/**54.0**/88.1	93.1/19.8/76.0	92.4/22.3/74.9
tile	96.4/55.8/90.6	98.9/94.6/98.0	98.9/92.0/97.1	**99.1**/**94.7**/**98.3**
toothbrush	98.9/39.2/91.5	99.0/70.7/94.0	**99.4**/**74.9**/**94.6**	**99.4**/74.7/94.2
transistor	**97.1**/**64.8**/**92.3**	86.4/45.5/74.2	79.1/31.3/63.3	76.6/27.9/66.3
wood	94.7/49.6/90.6	96.5/77.9/91.5	**97.6**/**83.9**/**94.2**	97.1/75.8/91.2
zipper	**99.0**/59.4/**96.4**	93.5/67.8/86.8	96.9/**74.7**/93.1	97.7/74.2/92.6
avg.	**98.4**/56.4/**94.6**	95.6/**67.5**/90.7	95.4/64.6/87.3	94.9/63.7/86.3

Table 3: Anomaly detection performance (image-level AUROC%) on the synthetic micro defect dataset. The best score among all methods for each object is indicated in bold.

Class	PatchCore	DRÆM	FIRE-AD	FIRE-AD$_{ag}$
bottle	**100**	92.2	94.1	80.9
cable	**87.7**	78.5	67.0	66.6
capsule	**100**	88.2	96.6	82.2
carpet	52.2	65.9	**68.3**	65.5
grid	93.6	**99.1**	89.9	77.2
hazelnut	**99.9**	94.4	98.6	92.7
leather	77.6	86.4	**92.6**	58.6
metal nut	**100**	97.1	98.8	98.2
pill	**100**	90.0	96.8	99.7
screw	**100**	98.5	99.7	99.8
tile	74.0	71.0	67.2	**74.1**
toothbrush	97.8	94.5	**98.6**	82.0
transistor	**99.9**	81.3	95.4	78.8
wood	75.8	**86.3**	60.4	60.7
zipper	**100**	86.4	**100**	90.8
avg.	**90.5**	86.0	88.2	80.5



(a) Input I (b) GT M (c) PatchCore $\widehat{M}$ (d) DRÆM $\widehat{M}$ (e) FIRE-AD $\widehat{M}$ (f) FIRE-AD$_{\text{ag}}$ $\widehat{M}$

Figure 4: Examples of input image I, ground-truth anomaly mask M, and estimated ones $\widehat{M}$ on MVTec AD dataset.

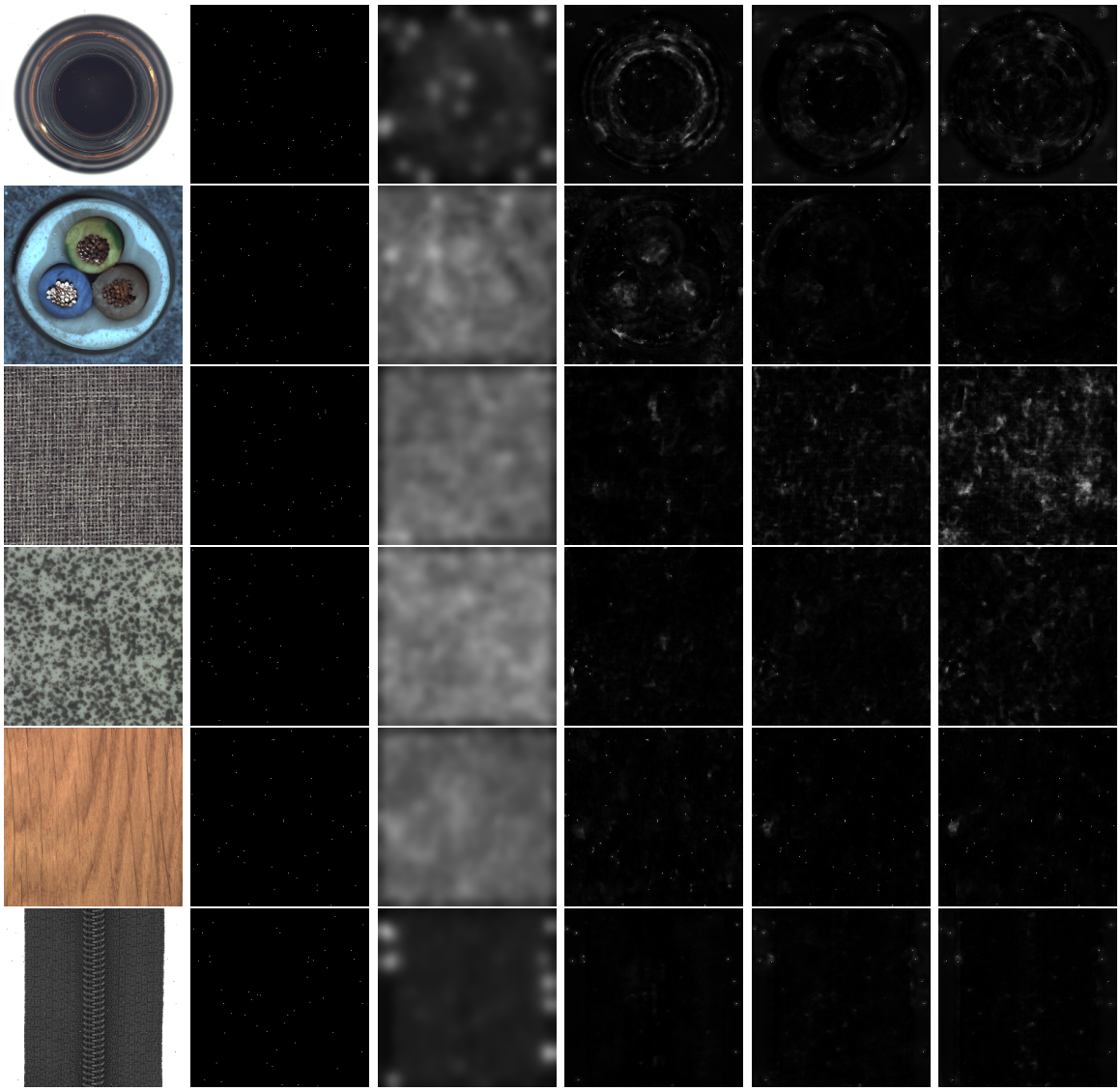


Figure 5: Examples of input image I, ground-truth anomaly mask M, and estimated ones $\widehat{M}$ on the micro defect dataset.

Table 4: Anomaly localization performance (pixel-level AUROC%/AP%) on the synthetic micro defect dataset. The best score among all methods for each object is indicated in bold.

Class	PatchCore	DRÆM	FIRE-AD	FIRE-AD$_{ag}$
bottle	86.6/0.3	**98.2**/76.7	97.8/**82.4**	97.6/81.9
cable	69.2/0.1	98.3/47.6	98.5/62.9	**98.7**/**73.5**
capsule	85.2/0.3	98.0/73.2	**98.8**/79.9	98.7/**80.0**
carpet	51.8/0.1	77.1/4.9	**78.4**/**7.4**	75.2/7.2
grid	67.3/0.1	**94.8**/69.9	94.2/69.0	94.2/**70.7**
hazelnut	84.6/0.3	**99.9**/96.4	**99.9**/97.0	**99.9**/**97.6**
leather	63.4/0.1	**99.9**/99.8	**99.9**/99.8	**99.9**/**99.9**
metal nut	82.8/0.3	97.7/84.2	**98.6**/86.9	**98.6**/**88.2**
pill	85.4/0.3	**98.9**/92.5	98.7/93.3	98.4/**94.1**
screw	84.3/0.3	94.0/66.5	96.5/69.6	**97.0**/**76.0**
tile	61.4/0.1	**89.4**/18.1	87.7/18.4	89.1/**18.6**
toothbrush	84.3/0.3	99.0/81.2	**99.3**/87.2	99.2/**88.1**
transistor	80.6/0.2	99.8/88.7	**99.9**/**92.4**	99.8/90.5
wood	64.7/0.1	**99.9**/99.8	**99.9**/**99.9**	99.9/99.8
zipper	78.3/0.2	88.2/29.0	89.7/29.3	**91.3**/**34.1**
avg.	75.3/0.2	95.5/68.6	**95.9**/71.7	95.8/**73.3**

5. Conclusion

In this work, we propose a new framework for detecting micro defects, called FIRE-AD, which separates images into multiple frequency domains and estimates an anomaly score map. This enables us to generate the anomaly score map by dynamic integration of reconstruction errors at each frequency domain. As a result of the evaluation on the MVTec AD and the synthetic micro defect datasets, we show the effectiveness of the proposed method over the existing method for detecting micro defects. Since the parameter indicating frequency band σ at the frequency separation network is fixed in this work, for further improvements, should be adaptively tuned to adjust to various-sized anomalies.

References

Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2019. doi: 10.5220/0007364503720380. URL https://doi.org/10.5220%2F0007364503720380.

Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4):1038–1059, 2021.

Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences, 2020. URL https://arxiv.org/abs/2005.02357.

Tamás Czimmermann, Gastone Ciuti, Mario Milazzo, Marcello Chiurazzi, Stefano Roccella, Calogero Maria Oddo, and Paolo Dario. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, 20(5):1459, 2020.

Naoyuki Ichimura. Spatial frequency loss for learning convolutional autoencoders, 2018. URL https://arxiv.org/abs/1806.02336.

Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection, 2022. URL https://arxiv.org/abs/2203.00259.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

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.

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 (ICCV)*, pages 8330–8339, October 2021.