Omni-Frequency Channel-Selection Representations for Unsupervised Anomaly Detection

Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, *Member, IEEE*, and Shuwen Pan

***Abstract*— Density-based and classification-based methods have ruled unsupervised anomaly detection in recent years, while reconstruction-based methods are rarely mentioned for the poor reconstruction ability and low performance. However, the latter requires *no costly extra training samples for the unsupervised training* that is more practical, so this paper focuses on improving reconstruction-based method and proposes a novel *O*mni-frequency *C*hannel-selection *R*econstruction (OCR-GAN) network to handle sensory anomaly detection task in a perspective of frequency. Concretely, we propose a Frequency Decoupling (FD) module to decouple the input image into different frequency components and model the reconstruction process as a combination of parallel omni-frequency image restorations, as we observe a significant difference in the frequency distribution of normal and abnormal images. Given the correlation among multiple frequencies, we further propose a Channel Selection (CS) module that performs frequency interaction among different encoders by adaptively selecting different channels. Abundant experiments demonstrate the effectiveness and superiority of our approach over different kinds of methods, *e.g.*, achieving a new state-of-the-art 98.3 detection AUC on the MVTec AD dataset without extra training data that markedly surpasses the reconstruction-based baseline by +38.1↑ and the current SOTA method by +0.3↑. The source code is available in the additional materials.**

***Index Terms*— Anomaly detection, omni-frequency decoupling, unsupervised learning, reconstruction-based network.**

I. Introduction

Anomaly detection is a binary classification task to distinguish whether a given image deviates from the predefined normality, which is an essential task in visual image understanding that has various applications in the real world, *e.g.*, novelty detection [1], product quality monitoring based on industrial images [2], automatic defect restoration [3], human health monitoring [4] and video surveillance [5], [6], [7], [8]. In real-world applications, anomaly detection tasks can be divided into sensory AD (Fig. 1a) and semantic AD (Fig. 1b): the former only suffers from covariate shift without semantic shift, while the later is the opposite. Most anomalies appear in the form of defects in the sensory AD, such as the normal defect detection task in MVTec AD [2] and KolektorSDD [9] datasets. However, semantic AD task detects images with label shifts, assuming that normal and abnormal come from different semantic distributions, such as the one-class detection task in CIFAR-10 [10]. This work focus on solving the sensory AD task but also evaluate on the related semantic AD dataset.

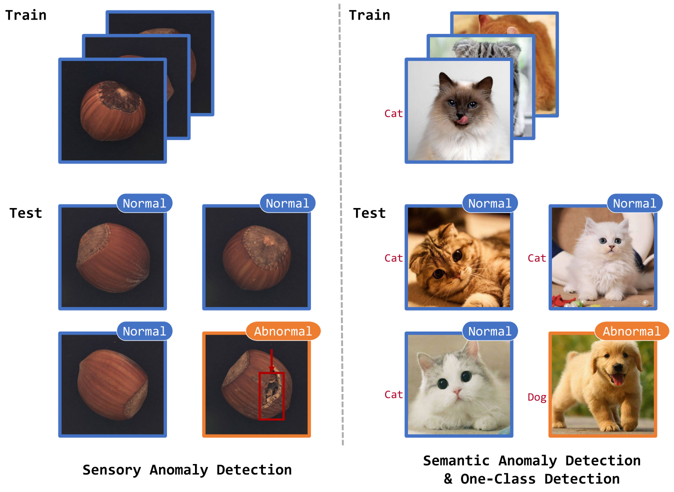


Fig. 1. Illustrations of sensory anomaly detection (**Left**) and semantic anomaly detection (**Right**).

In anomaly detection, obtaining abnormal samples and detecting novel abnormalities are time-consuming and costly objects that force us to develop unsupervised methods for more practical applications. Current unsupervised anomaly detection methods are mainly divided into three categories: density-based (Fig. 2a), classification-based (Fig. 2b) and reconstruction-based (Fig. 2c) methods. ***a)*** *Density-based methods* generally employ a pre-trained model to extract meaningful vectors of the input image. The anomaly score can be obtained by calculating the similarity between the embedding representation of the test image and the reference density distribution. This kind of method [11], [12], [13] achieves a high AUC score on the popular MVTec AD [2] dataset, but they *need pre-trained models and are insufficient for the model interpretability*. ***b)*** *Classification-based methods* try to find the

Manuscript received 15 March 2022; revised 11 March 2023; accepted 22 June 2023. Date of publication 24 July 2023; date of current version 2 August 2023. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant U21A20484 and in part by the Key Research and Development Program of Zhejiang Province under Grant 2022C01011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiangqian Wu. *(Yufei Liang and Jiangning Zhang contributed equally to this work.) (Corresponding author: Yong Liu.)*

Yufei Liang, Jiangning Zhang, and Yong Liu are with the Laboratory of Advanced Perception on Robotics and Intelligent Learning, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 22032139@zju.edu.cn; 186368@zju.edu.cn; yongliu@iipc.zju.edu.cn).

Shiwei Zhao and Runze Wu are with the NetEase Fuxi AI Laboratory, Hangzhou 310012, China (e-mail: zhaoshiwei@corp.netease.com; wurunze1@corp.netease.com).

Shuwen Pan is with the Discipline of Control Science and Engineering, School of Information and Electrical Engineering, Hangzhou City University, Hangzhou 310015, China (e-mail: pansw@zucc.edu.cn).

Digital Object Identifier 10.1109/TIP.2023.3293772

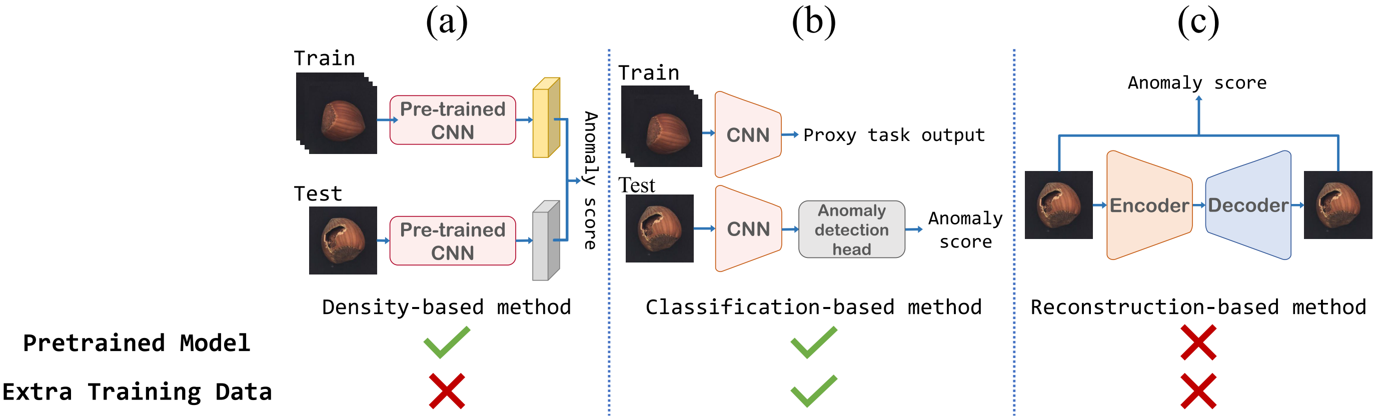
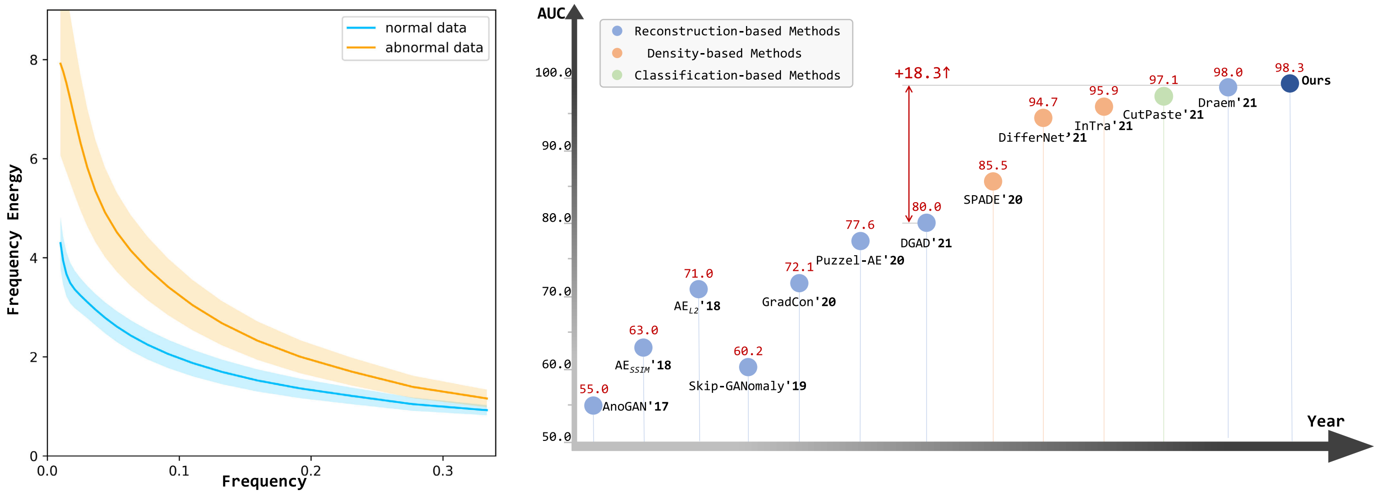


Fig. 2. Pipeline illustrations of three kinds of unsupervised anomaly detection methods in column. Bottom two rows indicate whether *Pretrained Model* and *Extra Training Data* are used for each kind of method.

classification boundaries of normal data. Self-supervised methods are representative of classification-based methods, which use the model trained by the proxy task to detect anomalies. Thus, self-supervised methods *rely on how well the proxy tasks match the test data*. For example, CutPaste [14] performs well in anomaly detection on MVTec AD dataset. However, it is difficult for this method to perform well on other datasets. Also, these methods *rely on pre-trained models and extra training data*. ***c)*** *Reconstruction-based methods* [15], [16], [17], [18] contain a generator to reconstruct the input image, and the anomaly score is the more interpretable reconstruction error. *These methods do not need pre-trained models and extra training data*. However, current reconstruction-based methods without extra training data are much less expressive than other methods. In summary, current unsupervised anomaly detection approaches are still suffering from two main challenges: ***(1)*** Some works achieve high AUC score but require abnormal samples or extra training data that are hard to obtain and costly for practical use. ***(2)*** Current reconstruction-based methods are more practical and do not need pre-trained models and extra training data but suffer from low performance. Although our approach borrows from self-supervised methods for constructing pseudo-anomaly data, this paper focuses on improving the reconstruction-based method as it requires no extra training data and only normal samples that is more practical.

To improve the performance of the reconstruction-based method, we need to enhance the reconstruction ability of the generator for the anomaly detection task. For an image, different frequency bands contain different types of information, *e.g.*, low frequency represents more semantic information while high frequency represents more detailed texture information. Also, we find that the model performance can be improved from the frequency domain perspective in many computer vision tasks, *e.g.*, in image super-resolution task, [20] separates the different frequency components to compensate for the loss of information in different frequency bands of real LR images to improve the performance of the model. Motivated by the idea, we analyze the frequency distribution of normal and abnormal images in the anomaly detection task. As shown in Fig. 3(a), we count the frequency energy distribution of normal and abnormal images, as the energy distribution of the Fourier-transformed image is reflected in the amplitude spectrum. We re-analyze this paradigm and find that normal and abnormal samples have different frequency distributions in sensory AD. So it may be difficult and unsuitable for only one generator to learn the full-frequency reconstruction of the RGB image. Therefore, we propose an anomaly detection framework using multiple frequency branches to reconstruct information from different frequency bands respectively. In order to differentiate the use of information from different frequency bands, we propose an effective *Frequency Decoupling* (FD) module to pre-obtain omni-frequency representation of the input image and use parallel generators to reconstruct images of multiple frequencies. Considering the model efficiency, we conduct experiments with 2 or 3 frequency branches in this paper. Different frequency branches in the framework are independent by default. However, an image contains information in multiple frequency bands, and the information in different frequency bands is not completely unrelated to each other but complementary in the real world. So, we design a tailored *Channel Selection* (CS) module to further realize omni-frequency interaction among multiple branches that can adaptively select different channel features. Based on the above modules and the baseline Skip-GANomaly [21], we propose a novel ***O****mni-frequency* ***C****hannel-selection* ***R****econstruction* (OCR-GAN) network. Our method achieves state-of-the-art (SOTA) results on multiple public datasets consistently. Specifically, our OCR-GAN improves +0.3↑ than current SOTA method Draem [19] and significantly +18.3↑ than SOTA reconstruction-based DGAD [17] without extra training data on MVTec AD in Fig. 3(b), emwhich strongly proves that the reconstruction-based method can also perform well even without extra training data and pre-trained models. To the best of our knowledge, this paper is the first attempt to explore omni-frequency information with reconstruction-based anomaly detection method. Our main contributions can be summarized as follows:

- We rethink the difference between normal and abnormal images from the frequency domain perspective and propose a novel framework for anomaly detection based on omni-frequency reconstruction.



(a) (b)

Fig. 3. (a) **Energy distribution** with frequencies for normal and abnormal samples in MVTec AD dataset, and the shadow represents standard deviation. *Normal and abnormal data have noticeable frequency distribution differences.* (b) **Development of three kinds of methods**. Our approach surpasses the SOTA reconstruction-based method without extra training data by a large margin, i.e., +18.3↑. Note that the current SOTA Draem [19] is not a classical reconstruction-based method, which requires a new training strategy and extra training data.

- We propose an effective FD module to obtain different frequency bands information of the image that enables the omni-frequency reconstruction by multiple branches.
- We propose a CS module to realize omni-frequency interaction among multiple branches and adaptive selection of different channel features.
- Abundant experiments demonstrate the superiority of our OCR-GAN over SOTA methods, *e.g.*, we achieve a new SOTA **98.3** detection AUC on the MVTec AD dataset without extra training data, which markedly surpasses the SOTA reconstruction-based method without extra training data by **+18.3**↑ and the SOTA method by **+0.3**↑.

The remainder of the paper is organized as follows. In Sec. II, we review some related works. Details of the proposed OCR-GAN method are given in Sec. III. Experimental results are presented in Sec. IV. And we conclude the paper with discussion and summary in Sec. V.

II. Related Work

Anomaly detection methods can be mainly divided into density-based, classification-based and reconstruction-based methods as follows.

A. Density-Based Methods

Density-based methods build a density estimation model for the distribution of normal training data. And this kind of method assumes that normal data have a higher likelihood under this model than abnormal data during inference. Parameter density estimation assumes that the density of normal data can be represented by some reference distribution. A pre-trained network is used to extract meaningful vectors representing the whole image or patch image for anomaly detection. The similarity between the representation vector of the test image and the reference vector is set as anomaly score. Some researches [22], [23], [24], [25], [26], [27] train the model on the entire image, while works [12], [13], [28], [29] on the patch image. The normal distribution reference can be the parameter of the Gaussian distribution of the normal image embedding vectors [13], [30], the mixed Gaussian distribution [31], [32], the Poisson distribution [33], the center of the sphere containing the embedding from normal images [12], [34], the entire set of normal embedding vectors [11], [35], the feature of the last layer in the network [36], [37], or the mid-level feature representation [38]. Mahalanobis distance is used to calculate the anomaly score between the embedding vector of the test image and the reference vector of the normal training distribution. The PaDIM [13] elaborate that the density-based model (*i.e.*, embedding similarity-based model) lacks interpretability. This method only performs anomaly detection and gives promising results. However, it lacks interpretability as it is impossible to know which part of an anomalous image is responsible for a high anomaly score. PaDIM interprets the location of anomalies by detecting them in the patch, but this introduces a large amount of computation. Also, this kind of method requires the pre-trained model for extracting vectors that is less practical for various real scenarios.

Another method of density estimation is normalizing flows. Normalizing flows are used to learn bijective transformations between data distributions with a special property. DifferNet [39] using normalizing flows to estimate the precise likelihood. Since flow-based methods have no dimensional reduction, the computation cost is significant. And this kind of method also needs pre-trained models to extract features.

B. Classification-Based Methods

Classification-based methods [40] try to find the classification boundaries of normal data. DeepSVDD [34] first introduces one-class classification to anomaly detection. Moreover, there are some self-supervised learning methods to design good proxy tasks to help the model detect anomalies from normal samples. One classical self-supervised anomaly detection method is isolation forest [41]. Other proxy tasks for self-supervised anomaly detection methods include image transformation prediction [24], [42], contrastive learning [43]

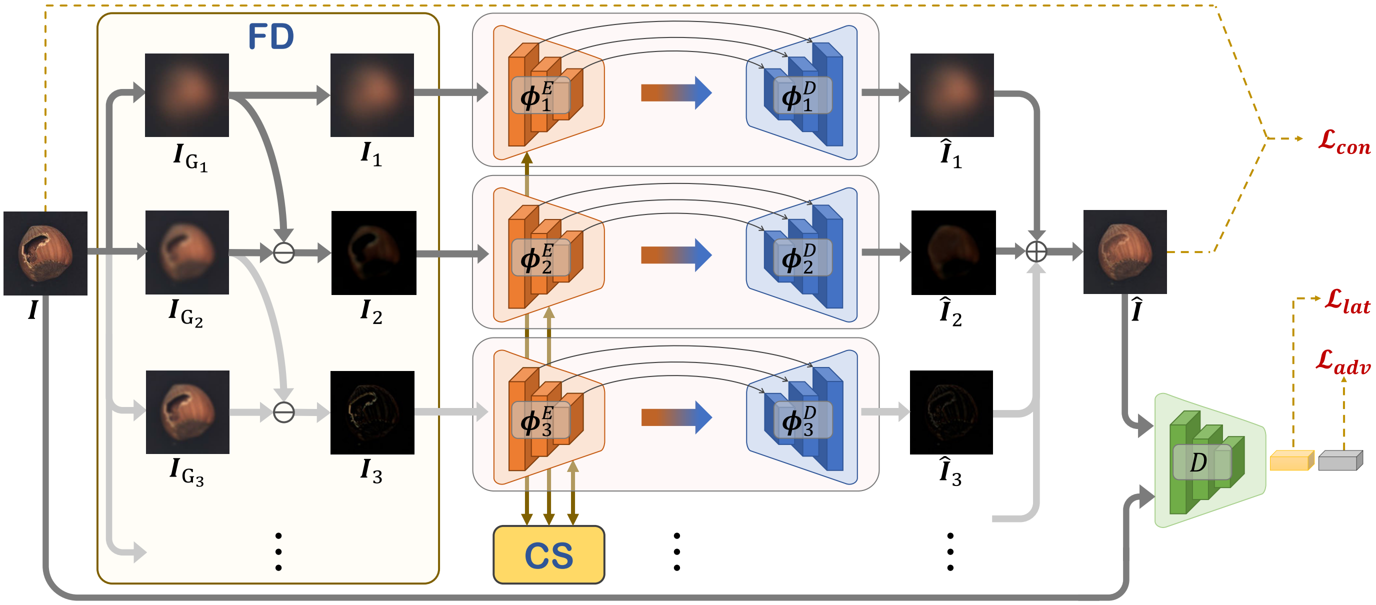


Fig. 4. **Overview of proposed OCR-GAN**. Input image $\boldsymbol{I}$ goes through Frequency Decoupling (FD) module to obtain omni-frequency images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots\}$ from pre-processed Gaussian images $\{\boldsymbol{I}_{G_1}, \boldsymbol{I}_{G_2}, \ldots\}$. Then $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots\}$ are fed into multiple generators $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots\}$ to reconstruct corresponding images $\{\hat{\boldsymbol{I}}_1, \hat{\boldsymbol{I}}_2, \ldots\}$, which are added to obtain the final output $\hat{\boldsymbol{I}}$. The proposed Channel Selection (CS) module performs omni-frequency interaction among different encoders, *i.e.*, $\{\boldsymbol{\phi}_1^E, \boldsymbol{\phi}_2^E, \ldots\}$.

and proxy binary classification [14]. Reference [14] uses data augmentation to generate pseudo-anomaly data and then does a binary classification proxy task with normal training samples to train the feature extraction model. The self-supervised method relies on the design of proxy tasks, which is difficult to perform well on multiple data sets. Representation learning performance in self-supervised methods relies on the ability of network feature extraction. Therefore, such methods usually use a pre-trained model as the feature extraction network.

C. Reconstruction-Based Methods

One of the reconstruction-based methods is sparse reconstruction which assumes that normal samples can be reconstructed with a limited number of basis functions while abnormal samples are more expensive to reconstruct. L_1 norm-based kernel PCA [44] and low-rank embedded networks [45] are belong to sparse reconstruction methods.

The reconstruction method is intuitive and easy to understand. Abnormal images would get higher reconstruction errors as they have a different data distribution than normal images. The autoencoder (AE) [46] and generative adversarial networks (GAN) [47] can reconstruct samples from the normal training data. [48] propose to use an autoencoder for the reconstruction process and structural similarity to measure reconstruction error. Some studies [47], [49] have shown that using adversarial network training would improve generation results. Moreover, GAN-based methods have more suitable metrics that can play the role of anomaly score, *e.g.*, output of the discriminator [50], [51] and latent space distance [21], [51], [52]. For more accurate anomaly detection, OCGAN [53] uses a denoising autoencoder, latent discriminator, visual discriminator, and classifier to ensure that any example generated from the learned latent space is indeed from the normal class. For GAN-based methods, the discriminator is usually used to distinguish the reconstructed image from the original image, but OGNet [54] redefines the role of the discriminator that is used to distinguish reconstructed images of different qualities. Recently, [55] utilize backpropagated gradients as representations to characterize anomalies, and The generation ability of the generator has a significant influence on the effect of the reconstruction-based method, so [56] propose to construct GAN ensembles for anomaly detection as GAN ensembles often outperform the single GAN. It is challenging to ensure poor reconstruction for abnormal samples as the capacity of the generator is strong. Thus these methods perform poorly in sensory detection. These methods indiscriminately reconstruct all frequencies of the RGB image that may be difficult for the generator, leading to poor results in anomaly detection. Also, we find that normal and abnormal samples have different frequency distributions, so we propose a new paradigm that uses parallel branches to reconstruct omni-frequency images.

III. Our Approach

A. Overview

In this section, we aim at improving the current reconstruction-based approach without extra training data and designing a generalized network for anomaly detection. As the difference between normal and abnormal images varies in different frequency bands, we perform anomaly detection from the perspective of the frequency domain. As shown in Fig. 4, our method derives from a frequency-decoupling idea that comprises multiple generators, *i.e.*, G=$\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots\}$, to reconstruct omni-frequency images $\{\hat{\boldsymbol{I}}_1, \hat{\boldsymbol{I}}_2, \ldots\}$, which is trained alternately with a discriminator D to further boost the model performance. Concretely, we propose an effective FD module to decouple the input image $\boldsymbol{I}$ to omni-frequency images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots\}$ and a CS module to realize omni-frequency interaction by adaptively selecting channels among encoders

$\{\phi_1^E, \phi_2^E, \dots\}$. When the model finishes the training, the abnormal images would be poorly reconstructed and get higher anomaly scores than normal images.

B. Frequency Decoupling

Pixel distributions reflect the spatial frequency of the image. Different frequency components contain different information, *e.g.*, the low frequency of the image contains more semantic information while the high frequency includes more details and texture information. As previously mentioned, normal and abnormal images have obvious frequency distribution differences, which derive from the abnormal elements in abnormal data, *e.g.*, holes, cracks, and scratches in the MVTec AD dataset. For a more thorough analysis, we counted the frequency energy distribution of normal and abnormal images. The energy distribution of the Fourier-transformed image is reflected in the amplitude spectrum. And Fig. 3(a) also shows the difference between normal and abnormal images in the frequency domain. Therefore, we consider that the importance of information in different frequency bands varies in anomaly detection tasks, especially sensory anomaly detection.

Motivated by the difference in the frequency distribution of normal and abnormal images shown in Fig. 3(a), we propose a tailored Frequency Decoupling (FD) module to pre-obtain informative omni-frequency representations. Specifically, FD contains the following three processes.

(1) Convolving original image $\boldsymbol{I}$ with the Gaussian kernel $\boldsymbol{Gau}_1$:

$$\boldsymbol{Gau}_1 = \frac{1}{256}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}, \tag{1}$$

and then removing even rows and columns of the blurred image to obtain intermediate down-sampled image $\boldsymbol{I}_{blur}$.

(2) $\boldsymbol{I}_{blur}$ would be exactly one-quarter the area of $\boldsymbol{I}$ and goes through a ×2↑ up-sampling to restore the original resolution, with the new even rows and columns filled with zeros. Then a similar convolution operation with the Gaussian kernel $\boldsymbol{Gau}_2 = 4 * \boldsymbol{Gau}_1$ is applied to approximate missing pixels, *i.e.*, zeros in even rows and columns, and we obtain first-level blurred image $\boldsymbol{I}_{G_{N-1}}$, where n represents the branch number (The smaller the N is, the less high-frequency information) and $\boldsymbol{I}_{G_N}$ is initialized with $\boldsymbol{I}$, denoted as:

$$\begin{aligned} \boldsymbol{I}_{G_N} &= \boldsymbol{I}, \\ \boldsymbol{I}_{G_{N-1}} &= \mathrm{Up}(\mathrm{Down}(\boldsymbol{I}_{G_N} * \boldsymbol{Gau}_1)) * \boldsymbol{Gau}_2, \end{aligned} \tag{2}$$

where $*$ means the convolution operation, while Down and Up are above-mentioned down-sampling and up-sampling operations. We would obtain a set of blurred images $\{\boldsymbol{I}_{G_1}, \boldsymbol{I}_{G_2}, \dots, \boldsymbol{I}_{G_n}\}$ by repeating the above processes.

(3) The blurred images $\boldsymbol{I}_{G_n}, n = 1, 2, \dots, N-1$ lost some high-frequency information in varying degrees, and we further calculate the difference between adjacent images to obtain omni-frequency images:

$$\begin{aligned} \boldsymbol{I}_1 &= \boldsymbol{I}_{G_1}, \\ \boldsymbol{I}_2 &= \boldsymbol{I}_{G_1} - \boldsymbol{I}_{G_2}, \\ &\vdots \\ \boldsymbol{I}_N &= \boldsymbol{I}_{G_{N-1}} - \boldsymbol{I}_{G_N}. \end{aligned} \tag{3}$$

FD can be effectively applied to frequency-sensitive tasks such as anomaly detection by decoupling different-frequency components as needed, and we set branch number two in the paper by default. Fig. 4 shows that different frequency components are reconstructed by multiple independent generators.

C. Channel Selection

Different frequency branches are relatively independent in our anomaly detection framework with only the FD module, which goes against the objective fact that different frequencies complement each other. Besides, as shown in Fig. 5 (a), the features of different channels in omni-frequency features are various. Attention [57] can help us to achieve the selection of information in different frequency bands. Therefore, we design a novel Channel Selection (CS) module to realize omni-frequency interaction among multiple branches and adaptive selection of different channel features. Fig. 5 shows the detailed structure of the CS module with a two-branch case that contains low and high-frequency features, but it is easy to extend to multiple branches. Concretely, for the given two feature maps $\boldsymbol{F}_h \in \mathbb{R}^{H\times W\times C}$ with high-frequency information and $\boldsymbol{F}_l \in \mathbb{R}^{H\times W\times C}$ with low-frequency information, we fuse them via an element-wise summation:

$$\boldsymbol{F} = \boldsymbol{F}_l + \boldsymbol{F}_h. \tag{4}$$

Then we apply Global Average Pooling to embed the global information and obtain channel-wise statistics $z^1 \in R^C$:

$$z_c^1 = \mathcal{F}_{GAP}(\boldsymbol{F}_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{F}_c(i,j), \tag{5}$$

where c is the c-th channel with C channels totally. After that, we use a fully connected layer to reduce the dimension of the embedded z^1 from C to d and obtain $z^2 \in R^d$, which is able to provide a precise and adaptive selection:

$$z^2 = \mathcal{F}_{FC}(z^1). \tag{6}$$

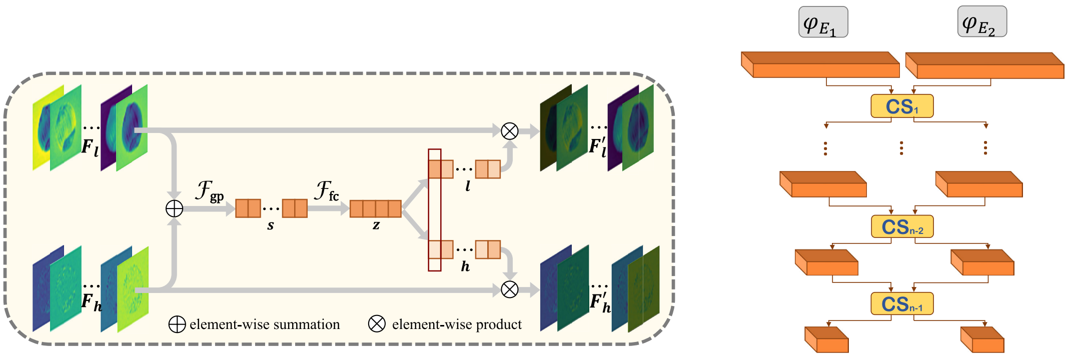
Finally, we use compact feature descriptor z^2 to regress c-th channel attentions for different frequency branch by:

$$\begin{aligned} \boldsymbol{l}_c &= \frac{e^{\boldsymbol{L}_c z^2}}{e^{\boldsymbol{L}_c z^2} + e^{\boldsymbol{H}_c z^2}}, \\ \boldsymbol{h}_c &= \frac{e^{\boldsymbol{H}_c z^2}}{e^{\boldsymbol{L}_c z^2} + e^{\boldsymbol{H}_c z^2}}, \end{aligned} \tag{7}$$

where $\boldsymbol{L}, \boldsymbol{H} \in R^{C\times d}$ represent the parameter weights, while $\boldsymbol{l}$ and $\boldsymbol{h}$ denote the channel attention vectors for $\boldsymbol{F_l}$ and $\boldsymbol{F_h}$. The augmented feature maps $\boldsymbol{F}'_l$ and $\boldsymbol{F}'_h$ are obtained through the channel attention operations as follows:

$$\begin{aligned} \boldsymbol{F}'_{l_c} &= \boldsymbol{l}_c \cdot \boldsymbol{F}_{l_c}, \\ \boldsymbol{F}'_{h_c} &= \boldsymbol{h}_c \cdot \boldsymbol{F}_{h_c}. \end{aligned} \tag{8}$$

CS module augments feature maps of different frequency branches by adaptively selecting channels, *i.e.*, using $\boldsymbol{l}$ and



(a) (b)

Fig. 5. **(a) Schematic diagram of CS**. For simplicity, a two-branch case is shown here that F_l and F_h represent low-/high-frequency features, *i.e.*, features in branch one and two. Augmented F'_l and F'_h are fed into following layers. **(b) Integration of CS to the framework**. The CS module is used at each stage of encoding.

$\boldsymbol{h}$ to re-weight $\boldsymbol{F}'_l$ and $\boldsymbol{F}'_h$. Note that the attention vectors of two branches complement each other, *i.e.*, $\boldsymbol{l}_c + \boldsymbol{h}_c = 1$, and this module can be easily extended to multiple branches by adding corresponding weights in Equation 7.

The CS module is applied to the encoding stage of the generator. We use different generators to encode different frequency band information. As shown in Fig. 11 (b), the feature maps of each layer of the individual encoders are used as input to the CS module, and the output of the CS module is used as input to the next layer of the individual encoders. In general, the CS module is used at each stage of encoding.

D. Training

Our OCR-GAN is trained from scratch end-to-end with only normal samples and tested with both normal and abnormal samples. We expect the GAN to correctly reconstruct normal samples both in image and latent vector space. Consistent with the previous reconstruction-based methods, our OCR-GAN assumes that out-of-distribution (*i.e.*, anomaly pixels) cannot be well reconstructed as the model is never trained on abnormal samples. Therefore, the difference between the reconstructed image and the input image in either image space or latent space is much greater for abnormal samples. Moreover, inspired by CutPaste [14], we also use data augmentation to generate forgery abnormal samples to assist the training process. Specifically, for normal images used in the training stage, we apply both CutPaste and CutOut [58] on each normal image to generate the forgery abnormal data. As shown in Fig. 6, the forgery abnormal samples generated by data augmentation and the samples generated by the generator are both used as positive inputs to the discriminator D, while original normal samples are used as negative inputs. As shown in Fig. 4, we adopt three losses for model training to ensure that OCR-GAN can well reconstruct normal samples.

1) Content Loss: The first term $\mathcal{L}_{con}$ ensures accurate reconstruction between the input normal image $\boldsymbol{I}$ and the reconstructed image $\hat{\boldsymbol{I}}$ by ℓ_1 error:

$$\mathcal{L}_{con} = \mathbb{E}_{\boldsymbol{I}\sim p_n}[\boldsymbol{I} - \hat{\boldsymbol{I}}]_1. \tag{9}$$

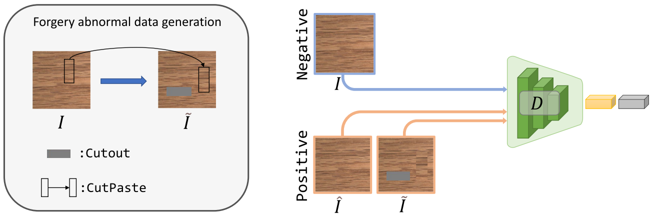


Fig. 6. Schematic diagram of using abnormal forgery samples to assist the training process.

This loss enables learning how to reconstruct similar images from the training data directly.

2) Adversarial Loss: The second term $\mathcal{L}_{adv}$ employs a discriminator D for adversarial training [47], which significantly improves the quality of the constructed image. Our model is to minimize it for G and maximize for D. Adversarial loss allows the generator to reconstruct the image as realistically as possible, while the discriminator distinguishes the normal images from the reconstructed and forgery abnormal images:

$$\mathcal{L}_{adv} = \mathbb{E}_{\hat{\boldsymbol{I}}\sim p_r}[D(\hat{\boldsymbol{I}})] + \mathbb{E}_{\tilde{\boldsymbol{I}}\sim p_f}[D(\tilde{\boldsymbol{I}})] - \mathbb{E}_{\boldsymbol{I}\sim p_n}[D(\boldsymbol{I})]. \tag{10}$$

3) Latent Loss: Latent loss [21] penalizes the similarity between the positive and negative images in the latent space. In OCR-GAN, we use the features of the last convolutional layer of the discriminator D as latent space features. Then, the ℓ_2 error between latent space features is used as the latent loss:

$$\mathcal{L}_{lat} = \mathbb{E}_{\boldsymbol{I}\sim p_n}[D_{lat}(\boldsymbol{I}) - D_{lat}(\hat{\boldsymbol{I}})]_2. \tag{11}$$

Note that $\hat{\boldsymbol{I}} = G(\boldsymbol{I})$, $\tilde{\boldsymbol{I}}$ is the forged abnormal data obtained by data augmentation, p_n, p_r and p_f are normal, reconstructed and forged image distributions, and $D_{lat}(\cdot)$ denotes the feature extraction of D for the penultimate layer. The total loss $\mathcal{L}_{all}$ is a weighted sum of above losses:

$$\mathcal{L}_{all} = \lambda_{con}\mathcal{L}_{con} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{lat}\mathcal{L}_{lat}. \tag{12}$$

Based on the baseline skip-GANomaly settings and the results of our experiments, the weight parameters are chosen as $\lambda_{rec} = 50$, $\lambda_{adv} = 1$, and $\lambda_{lat} = 1$.

E. Inference

Anomaly score proposed in [59] is used to detect anomalies during inference. For a test image $\boldsymbol{I}$, its anomaly score is defined as:

$$A(\boldsymbol{I}) = \lambda \mathcal{L}_{con}(\boldsymbol{I}) + (1 - \lambda)\mathcal{L}_{lat}(\boldsymbol{I}), \tag{13}$$

where $\mathcal{L}_{con}(\boldsymbol{I})$ is the reconstruction error that measures the content similarity between the input and reconstructed images, while $\mathcal{L}_{lat}(\boldsymbol{I})$ is the latent representation error based on the latent loss. Following the setting of our baseline skip-GANomaly, the weight parameter λ is set to 0.9.

Based on Equation 13, we are able to compute the anomaly score for each test sample in the test set. The set of anomaly scores for all samples in the test set is $\boldsymbol{A}$. Following [21], we scale $\boldsymbol{A}$ to [0, 1]. Therefore, the final anomaly score for a test image $\boldsymbol{I}$ is:

$$A'(\boldsymbol{I}) = \frac{A(\boldsymbol{I}) - min(\boldsymbol{A})}{max(\boldsymbol{A}) - min(\boldsymbol{A})}. \tag{14}$$

Threshold of anomaly score can be set according to requirements in real world applications.

IV. Experiments

In order to assess the effectiveness of the proposed OCR-GAN, we consider two types of anomaly detection tasks: sensory AD and semantic AD. We evaluate the performance of the proposed OCR-GAN in two cases against state-of-the-art methods using public-available datasets.

A. Experimental Setup

1) Datasets: This paper focuses on the sensory AD task. Thus, we use MVTec AD [2], DAGM [60] and KolektorSDD [9] to evaluate the performance of OCR-GAN in sensory AD. To further validate that OCR-GAN can improve the generation ability of generators in anomaly detection tasks, we use CIFAR-10 [10] to evaluate the performance of OCR-GAN in semantic AD.

MVTec AD [2] contains 5,354 high-resolution color images that consist of 10 kinds of objects and 5 kinds of textures, which is widely used for the anomaly detection task. The image resolution ranges from 700 to 1,024, and we downscale all images to 256 × 256 resolution for all experiments. The number of training samples for each category ranges from 60 to 320, and the abnormal samples in the test set contain more than 70 defects, *e.g.*, cracks, scratches, deformation, and holes.

DAGM [60] is a well-known benchmark database for surface defect detection. It contains images of various surfaces with artificially generated defects. Surfaces and defects are split into 10 classes of various difficulties. It is a weakly supervised dataset, and there are 8,050 training and testing sets each, and the ratio of positive and negative samples for each type is approximately 1:7. OCR-GAN is trained only on anomaly-free training samples for all experiments.

KolektorSDD [10] is constructed from images of defected electrical commutators. Specifically, microscopic fractions or cracks are observed on the surface of the plastic embedding in electrical commutators. The dataset contains 50 commutator samples, each with 8 surfaces, totaling 399 images in 500×1.240, of which 347 images are without any defect, and 52 images are with visible defects. The anomalies are tiny and visually similar to the background, making this dataset challenging for anomaly detection.

CIFAR-10 [10] consists of 60,000 color images in 32×32 with 10 classes. Semantic AD experiments regard one class as normal and the other classes as abnormal. CIFAR-10 is a challenging semantic AD dataset because images differ substantially across classes, and the background of images are not aligned. We experimented under two settings. Setting1 (S1) is protocol 2 described in [53], which uses the whole training set of just one class as the normal data for training and the whole test set for the test time. Setting2 (S2) is the setting used in skip-GANomaly, that CIFAR-10 can yield 10 different anomaly cases, each with 45,000 normal training samples and 9,000:6,000 normal-anomaly test samples.

2) Implementation Details: Our method is implemented by PyTorch 1.2.0 [65] and CUDA 10.2, and all experiments run with a TITAN RTX GPU. We use Adam [66] optimizer and set $\beta_1 = 0.5$, $\beta_2 = 0.999$, weight-decay=$1e^{-4}$, and learning rate=0.002. Unless otherwise specified, the batchsize is set to 32 for MVTec AD dataset, 64 for DAGM dataset, 64 for KolektorSDD dataset and 64 for CIFAR-10 dataset, and we use two frequencies (denoted as low-/high-frequency branches) for experiments. We choose Skip-GANomaly [21] as our baseline.

3) Evaluation Metrics: The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) Curve is used as a standard evaluation metric for anomaly detection. It is calculated by gradually changing the threshold of anomaly scores. AUC is accumulated to a score for the performance evaluation. A higher AUC score means better anomaly detection performance.

4) Default Setting: Following the previous unsupervised anomaly detection methods settings, our ablation experiments and interpretability experiments are all conducted on the MVTec AD dataset unless otherwise specified. Considering the number of parameters and the computational cost, we choose two frequency branches (high frequency and low frequency) for the experiments if not specified. Moreover, our OCR-GAN keeps the same parameter settings as the baseline. If not specified, the number of feature channels of each generator is set to 64.

B. Compare With SOTA Methods

We evaluate the performance of our OCR-GAN on four popular datasets to verify the superiority of our method over other SOTA methods.

1) Sensory AD: The difference between normal and abnormal images in the sensory AD task is covariate shift. And anomalies usually appear in the form of defects, such as cracks, scratches and holes. Actually, we design our OCR-GAN based on the motivation of difference in the frequency distribution of normal and abnormal samples in the MVTec AD dataset (shown in Fig. 3(a)). And the frequency analysis can be extended to other sensory AD datasets. Therefore, we choose three different sensory AD datasets to

TABLE I

AUC Results With SOTAs on MVTec AD Dataset. Three to Ten Columns Are Reconstruction-Based Methods While the Following Four Columns Are Density-Based and Classification-Based Methods. Bold and underline Represent Optimal and Suboptimal Results. ’ Means the Year of Publication. † Means Using the Pre-Trained Model With Extra Dataset. ‡ Means Our Training With Forgery Abnormal Samples in Sec. IV-C

	Items	AGAN	AE_1	AE_2	SkipG	GradC	P-AE	DGAD	Draem	Diff	CutPaste	CutPaste†	InTra	Ours	Ours‡
		[59]	[48]	[48]	[21]	[55]	[61]	[17]	[19]	[39]	[14]	[14]	[62]		
		17’	18’	18’	19’	20’	20’	21’	21’	21’	21’	21’	21’		
texture	Carpet	49.0	67.0	50.0	70.9	89.3	65.7	52.0	97.0	92.9	93.1	**100.0**	98.8	$98.9_{\pm 0.5}$	$\underline{99.4}_{\pm 0.3}$
	Grid	51.0	69.0	78.0	47.7	71.6	75.4	67.0	$\underline{99.9}$	84.0	$\underline{99.9}$	99.1	**100.0**	$99.6_{\pm 0.2}$	$99.6_{\pm 0.2}$
	Leather	52.0	46.0	44.0	60.9	69.3	72.9	94.0	**100.0**	$\underline{97.1}$	**100.0**	**100.0**	**100.0**	$\underline{97.1}_{\pm 0.6}$	$\underline{97.1}_{\pm 0.8}$
	Tile	51.0	52.0	77.0	29.9	63.4	65.5	83.0	$\underline{99.6}$	99.4	93.4	**99.8**	98.2	$92.2_{\pm 0.8}$	$95.5_{\pm 1.5}$
	Wood	68.0	83.0	74.0	19.9	76.7	89.5	72.0	$\underline{99.1}$	**99.8**	98.6	**99.8**	98.0	$95.8_{\pm 1.6}$	$95.7_{\pm 1.1}$
	Average	54.2	63.4	64.6	45.86	74.1	73.8	73.6	$\underline{99.1}$	94.6	97.0	**99.7**	99.0	$96.6_{\pm 0.3}$	$97.5_{\pm 0.3}$
object	Bottle	69.0	88.0	80.0	85.2	52.0	94.2	97.0	99.2	99.0	98.3	**100.0**	**100.0**	$\underline{99.6}_{\pm 0.2}$	$\underline{99.6}_{\pm 0.1}$
	Cable	53.0	61.0	56.0	54.4	58.7	87.9	90.0	91.8	95.9	80.6	96.2	84.2	$\mathbf{99.2}_{\pm 0.5}$	$\underline{99.1}_{\pm 0.6}$
	Capsule	58.0	61.0	62.0	54.3	55.6	66.9	60.0	**98.5**	86.9	$\underline{96.2}$	95.4	86.5	$95.4_{\pm 0.4}$	$\underline{96.2}_{\pm 0.6}$
	Hazelnut	50.0	54.0	88.0	24.5	91.4	91.2	80.0	**100.0**	99.3	97.3	$\underline{99.9}$	95.7	$88.2_{\pm 2.0}$	$98.5_{\pm 1.3}$
	Metal Nut	50.0	54.0	73.0	81.4	56.0	66.3	95.0	98.7	96.1	$\underline{99.3}$	98.6	96.9	$98.7_{\pm 0.2}$	$\mathbf{99.5}_{\pm 0.3}$
	Pill	62.0	60.0	62.0	67.1	92.4	71.6	76.0	**98.9**	88.8	64.7	93.3	90.2	$\underline{98.5}_{\pm 0.4}$	$98.3_{\pm 0.2}$
	Screw	35.0	51.0	69.0	87.9	78.2	57.8	67.0	93.9	96.3	86.3	86.6	95.7	$\mathbf{100.0}_{\pm 0.0}$	$\mathbf{100.0}_{\pm 0.0}$
	Toothbrush	57.0	74.0	98.0	58.6	98.0	97.8	93.0	**100.0**	98.6	98.3	90.7	**99.7**	$98.2_{\pm 0.9}$	$98.7_{\pm 0.7}$
	Transistor	67.0	52.0	71.0	84.5	72.8	86.0	88.0	93.1	91.1	95.5	$\underline{97.5}$	95.8	$94.9_{\pm 0.3}$	$\mathbf{98.3}_{\pm 1.5}$
	Zipper	59.0	80.0	80.0	76.1	56.6	75.7	82.0	**100**	95.1	99.4	$\underline{99.9}$	99.4	$97.6_{\pm 0.4}$	$99.0_{\pm 0.2}$
	Average	56.0	63.5	73.9	67.4	71.2	79.5	82.8	$\underline{97.4}$	94.7	94.3	95.8	94.4	$97.0_{\pm 0.2}$	$\mathbf{98.7}_{\pm 0.3}$
	All	55.0	63.0	71.0	60.2	72.1	77.6	80.0	$\underline{98.0}$	94.7	95.2	97.1	95.9	$96.9_{\pm 0.2}$	$\mathbf{98.3}_{\pm 0.2}$

evaluate the effectiveness of our method in the sensory AD task.

2) MVTec AD: Tab. I shows the detection AUC results of different methods on the MVTec AD dataset, and our experiments run five times using different random seeds without extra training data. We report the mean AUC score with corresponding standard error for each category and the average AUC for texture, object, and all categories. Results indicate that our approach achieves a new SOTA on the MVTec AD dataset without extra training data, *i.e.*, obtaining 98.3 detection AUC score. OCR-GAN improves the AUC score by a significant +18.3↑ compared with SOTA classical reconstruction-based method DGAD [17] without extra training data, by 3.6↑ compared with SOTA density-based method DifferNet [39], and by 1.2↑ compared with SOTA classification-based method CutPaste [14]. Our OCR-GAN belongs to classical reconstruction-based approaches that use the generator to reconstruct the image and use the reconstruction error to detect anomalies. The current SOTA method, Draem [19], trains two models (reconstruction model and anomaly segmentation model) using extra training data and uses the results of the anomaly segmentation to detect anomalies, which is very different from the classical reconstruction-based approaches. However, our OCR-GAN achieves a higher AUC score than Draem [19] by +0.3 ↑ without using extra training data, proving the effectiveness of our approach. Although our OCR-GAN does not achieve the best performance in every category, we obtain the highest overall score, and the AUC score of each category surpasses 95 (*c.f* Ours‡ in the table) proves the robustness and practicality of our method. In conclusion, it is remarkable that we firstly achieve SOTA on the MVTec AD dataset with a classical reconstruction-based method without extra training data while previous classical reconstruction-based methods fail to perform so well on sensory AD.

3) DAGM: Our OCR-GAN is trained only on normal samples for the DAGM dataset. Experimental results of other unsupervised methods on DAGM are obtained by experimenting using their open-source code. Tab. II shows that supervised methods have achieved near-perfect AUC on the DAGM dataset. Compared with supervised methods, previous unsupervised methods (including classification-based methods, reconstruction-based methods, and density-based methods) did not perform well on this dataset. However, our OCR-GAN achieves a 99.3 detection AUC score without extra training data. Our performance is comparable to supervised methods, which is a remarkable result.

4) KolektorSDD: OCR-GAN is compared with other unsupervised methods, and results are shown in Tab. III. As the anomaly elements in the dataset are small and similar to the background, previous unsupervised methods do not perform well on the KolektorSDD dataset. Without extra training data, our OCR-GAN achieves a 91.4 detection AUC score that increases by +5.5↑ over the suboptimal method Draem which is trained with extra data. This result shows that our approach also performs reliably in challenging sensory AD dataset.

5) Semantic AD: Experiments on the semantic anomaly detection task are performed to assist in proving the effectiveness of our approach. The difference between normal and abnormal images in the semantic AD [67] task is label shift. There is no variability across frequency bands of normal and abnormal in this task. However, our method can also improve the generation ability of the generator. We choose a public dataset CIFAR-10, which is widely used for one-class

TABLE II

AUC Results With SOTAs on DAGM Dataset. Bold and underline Represent Optimal and Suboptimal Unsupervised Results. *: Original Paper Only Reports Average AUC, and the Corresponding Line Shows Our Reproduced Results

	Methods	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10	Average
Unsup.	skipGAN [21]	58.3	56.1	55.1	53.7	57.4	66.8	52.4	53.7	52.3	52.2	55.8
	Puzzle AE [61]	50.7	50.5	58.7	70.0	63.6	92.3	54.0	49.1	54.6	49.6	59.3
	CutPaste [14]	56.1	87.8	57.1	71.3	47.4	68.8	96.5	53.4	51.9	74.7	66.0
	DifferNet [39]	59.7	82.9	69.8	97.3	61.2	97.0	68.5	52.1	78.2	79.1	74.6
	Draem [19]	96.1	98.3	**99.5**	**99.6**	92.1	**100**	99.7	**99.9**	98.9	96.0	98.0 (99.0*)
	Ours	**99.1**	**100**	99.1	99	**100**	97.5	**99.8**	99.8	**99.5**	**99.2**	**99.3**
Sup.	Lin *et al.* [63]	100	94.0	100	100	100	100	100	99.0	100	100	99.3
	Bǒ *et al.* [64]	100	100	100	100	99.9	100	100	100	100	100	100

TABLE III

AUC Results With SOTAs on KolektorSDD Dataset. Bold and underline Represent Optimal and Suboptimal Results

Methods	AUC
skipGAN [21]	55.1
Puzzle AE [61]	55.4
DifferNet [39]	84.9
InTra [62]	70.1
CutPaste [14]	60.2
Draem [19]	85.9
Ours	**91.4**

detection (semantic AD), to verify the effectiveness of our method on all types of anomaly detection tasks.

6) CIFAR-10: As shown in Tab. IV, reconstruction-based methods perform better in the semantic AD compared to the density-based methods and the classification-based methods. 1) In the experiment under Setting1, our OCR-GAN outperforms all compared unsupervised methods and achieves 79.5 AUC on the CIFAR-10 dataset, which is +7.0↑ higher than the suboptimal method. 2) In the experiment under Setting2, the performance of our method is greatly improved +16.3↑ compared to the baseline skip-GANanomaly. The results verify that our OCR-GAN performs well in one-class detection (semantic AD). Although the difference in frequency bands is not suitable for semantic anomaly detection, the improvement to structural design implicitly improves the model's reconstruction ability. Therefore the performance of the model in semantic AD is also enhanced.

C. Ablation Study

1) Influence of Different Components: We further conduct an ablation study on the MVTec AD dataset to investigate the effectiveness of each component of the proposed OCR-GAN. We choose Skip-GANomaly [21] as our baseline and gradually add different component, performing following seven experiments: ***(1)*** Baseline; ***(2)*** Baseline training with forgery abnormal images by both cutout and cutpaste data augmentations; ***(3)*** Adding FD module; ***(4)*** (3) training with forgery abnormal images by both cutout and cutpaste data augmentations; ***(5)*** Adding both FD and CS modules, *i.e.*, OCR-GAN in the paper; ***(6)*** Using three frequency branches; ***(7)*** OCR-GAN training with forgery abnormal images by the cutout data augmentation; ***(8)*** OCR-GAN training with forgery abnormal images by the cutpaste data augmentation; ***(9)*** OCR-GAN training with forgery abnormal images by both cutout and cutpaste data augmentations. As shown in Tab. V, our baseline only obtains a 60.2 AUC score because this reconstruction-based method suffers from the poor reconstruction ability of the generator. Training with forgery abnormal images can bring +8.3↑ AUC improvement to the baseline model. When the FD module is added to the baseline, the model performance increases by a significant +14.6↑, and our proposed CS module further improves the AUC score by +22.1↑ to 96.9. And adding these two modules to the baseline with cutout/cutpaste still gives a significant improvement in anomaly detection performance, which verifies that the main improvement of the model is from FD and CS. The results strongly demonstrate the effectiveness of our proposed two modules for the anomaly detection task. Moreover, our approach improves by +0.6↑ when using three frequency branches, meaning that more frequencies contribute to the model performance. We set the frequency number as two in the paper to balance model effectiveness and efficiency. And we use data augmentation to generate forgery abnormal samples to assist the training process. As shown in Tab. V, each data augmentation contributes to the model performance, and our OCR-GAN obtains the best result when both augmentations are applied.

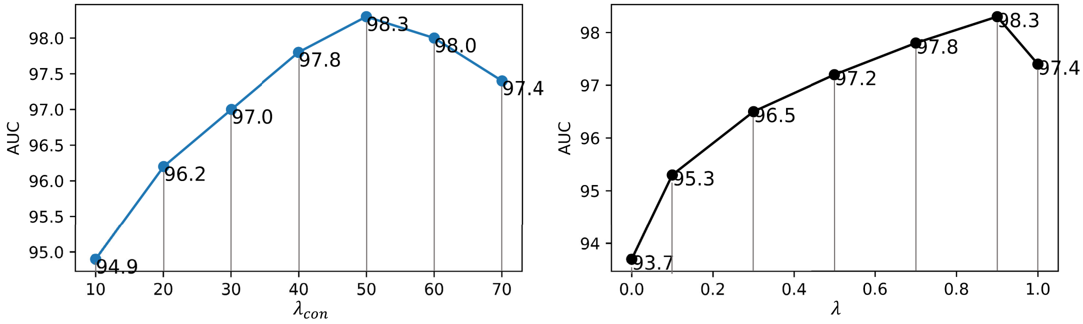
2) Influence of Frequency Branches: Flat distribution corresponds to the low-frequency component, while sharp changes (e.g., edge and noise) denote the high-frequency term. As different frequencies contain different information, we conduct an ablation study on the MVTec AD dataset to explore the anomaly detection performance using different frequency branches. As shown in Tab. VI, we conduct a set of experiments using only a high-frequency branch, only a low-frequency branch, two independent frequency branches, and two frequency branches with CS module (two-frequency-branch OCR-GAN). Results show that using only high-frequency information performs better than low-frequency information, meaning that abnormal elements contain more high-frequency information. Nevertheless, using two-frequency branches independently is not ideal for lacking the information interaction between different frequency branches. Our designed CS module can well handle this problem and further improve the model performance.

3) Influence of Hyperparameter Values: The training and testing process of our OCR-GAN involves the choosing of

TABLE IV

AUC RESULTS WITH SOTAS ON CIFAR-10 DATASET. BOLD AND UNDERLINE REPRESENT OPTIMAL AND SUBOPTIMAL RESULTS

	Methods	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Average
S1	OCSVM [36]	63.0	44.0	64.9	48.7	73.5	50.0	72.5	53.3	64.9	50.8	58.6
	AnoGAN [59]	67.1	54.7	52.9	54.5	65.1	60.3	58.5	62.5	75.8	66.5	61.8
	skipGAN [21]	65.6	47.6	66.0	57.8	74.6	58.7	61.6	64.7	76.1	69.1	64.2
	OCGAN [53]	75.7	53.1	64.0	62.0	72.3	62.0	72.3	57.5	82.0	55.4	65.7
	Gradcon [55]	76.0	59.8	64.8	58.6	73.3	60.3	68.4	56.7	78.4	67.8	66.4
	Puzzle AE [61]	78.9	78.0	**70.0**	54.9	75.5	66.0	74.8	73.3	83.3	70.0	72.5
	Draem [19]	58.8	56.5	55.6	58.5	53.0	64.7	59.0	54.3	51.0	54.4	56.6
	CutPaste [14]	70.0	62.0	62.6	62.0	53.8	62.9	62.4	59.9	51.8	57.6	60.5
	Intra [62]	50.2	48.9	57.8	49.2	55.4	60.3	44.5	65.7	73.8	64.9	57.1
	Ours	**82.0**	**78.9**	68.4	**78.9**	**84.5**	**80.8**	**75.1**	**89.6**	**84.4**	**72.2**	**79.5**
S2	skipGAN [21]	44.8	**95.3**	60.7	60.2	61.5	**93.1**	78.8	**79.7**	65.9	**90.7**	73.1
	Ours	**99.9**	80.2	**82.5**	**85.4**	**98.5**	87.3	**98.6**	76.9	**99.8**	85.2	**89.4**



(a)

(b)

Fig. 7. (a): Ablation study for the value of λ_{con} in loss function; (b): Ablation study for the value of lambda in anomaly score.

TABLE V

ABLATION STUDY ON MVTEC AD DATASET. BN REPRESENTS BRANCH NUMBER

No.	BN	FD	CS	cutout	cutpaste	AUC
(1)	2	✗	✗	✗	✗	60.2
(2)	2	✗	✗	✓	✓	$68.5_{+8.3}$
(3)	2	✓	✗	✗	✗	$74.8_{+14.6}$
(4)	2	✓	✗	✓	✓	$77.2_{+17.0}$
(5)	2	✓	✓	✗	✗	$96.9_{+36.7}$
(6)	3	✓	✓	✗	✗	$97.5_{+37.3}$
(7)	2	✓	✓	✓	✗	$97.4_{+37.2}$
(8)	2	✓	✓	✗	✓	$97.5_{+37.3}$
(9)	2	✓	✓	✓	✓	$98.3_{+38.1}$

TABLE VI

ABLATION STUDY FOR FREQUENCY BRANCHES

Category	high frequency	low frequency	two branches	OCR-GAN
texture	85.2	69.8	73.6	96.6
object	81.3	75.1	75.4	97.0
all	82.6	73.3	74.8	96.9

TABLE VII

ABLATION STUDY OF USING DIFFERENT LOSS COMPONENTS

$\mathcal{L}_{con}$	$\mathcal{L}_{adv}$	$\mathcal{L}_{lat}$	AUC
✓	✗	✗	96.3
✗	✓	✗	90.5
✗	✗	✓	94.5
✓	✓	✗	97.1
✗	✓	✓	95.0
✓	✗	✓	97.6
✓	✓	✓	98.3

several hyperparameters, such as weighting parameters for different losses in Equation(12) and weighting parameters for reconstructed differences and latent space differences in Equation(13). The choice of these parameters in our experiments follows the parameter settings of our baseline skip-GANomaly. We further conduct experiments to explore the influence of the value of the hyperparameters on the anomaly detection performance of the model. We conduct the ablation experiments about the parameter settings in Equation(11). Fig. 7(a) shows how the value of λ_{con} in the loss function influences the model performance, and Tab. VII shows the ablation study of using different loss components. The results show the contribution of each loss component to the anomaly detection performance. We also conduct experiments to explore the influence of the value of lambda in Equation(13). As shown in Fig. 7(b), when this parameter is set to 0.9, the model achieves the best anomaly detection performance.

4) Influence of Number of Parameters: Our OCR-GAN reconstructs the information of different frequency bands

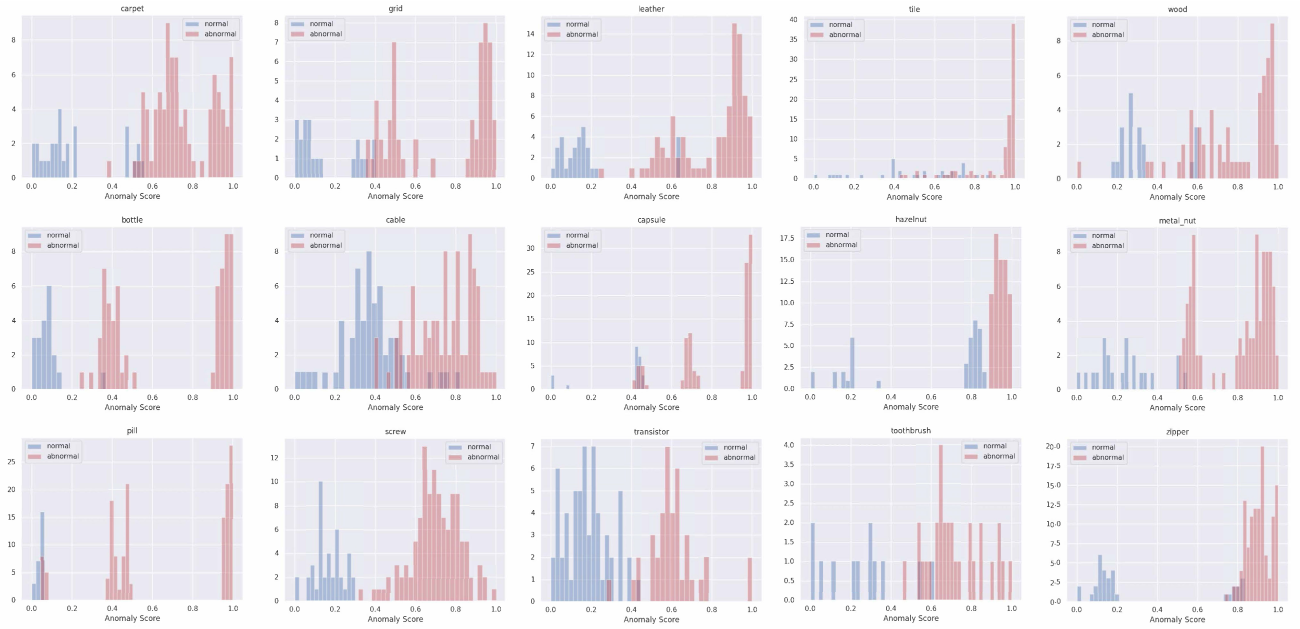
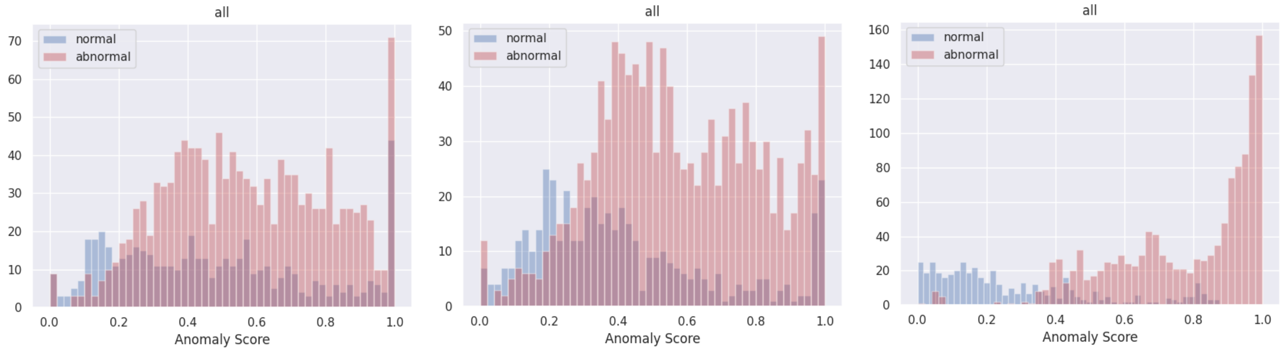


Fig. 8. Histogram of anomaly scores for the normal and abnormal samples for each category in the MVTec AD dataset.



(a) (b) (c)

Fig. 9. Comparison of anomaly score histograms for all category. **(a)**:Baseline. **(b)**:Adding FD. **(c)**:Adding both FD and CS.

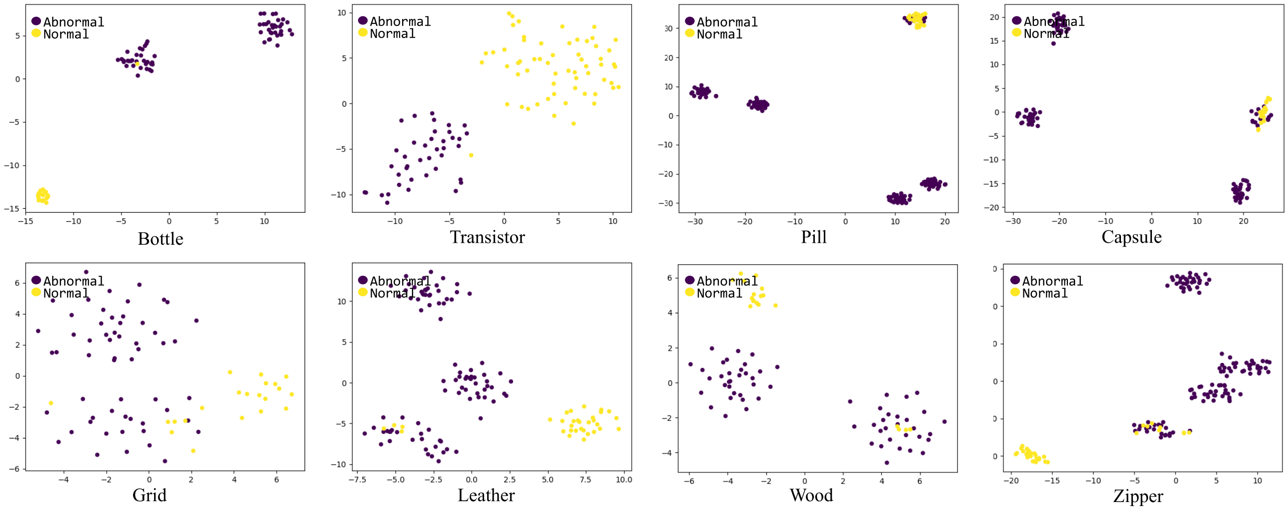


Fig. 10. **t-SNE visualization** of normal and abnormal samples for eight categories in MVTec AD dataset.

respectively, which means that one frequency band requires one generator. Considering the model efficiency, we conduct experiments mostly with two frequency branches. Compared with baseline, our OCR-GAN has a larger number of method parameters. So, we reduce the number of feature channels of the generator to explore the influence of model parameters

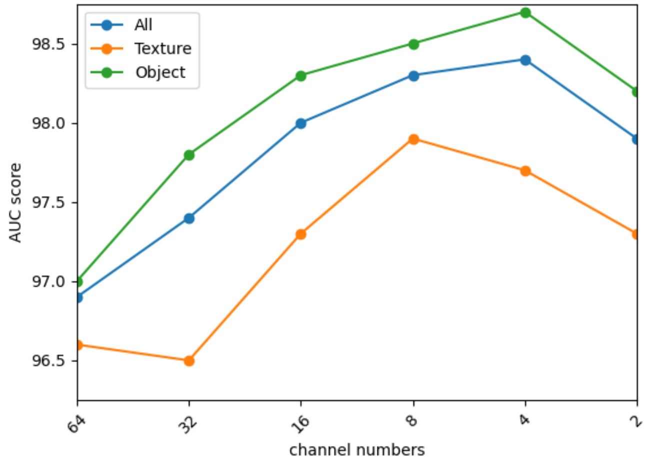


Fig. 11. **Influence of channel numbers.** With the number of generator feature channels decreases, the AUC first increases and then decreases.

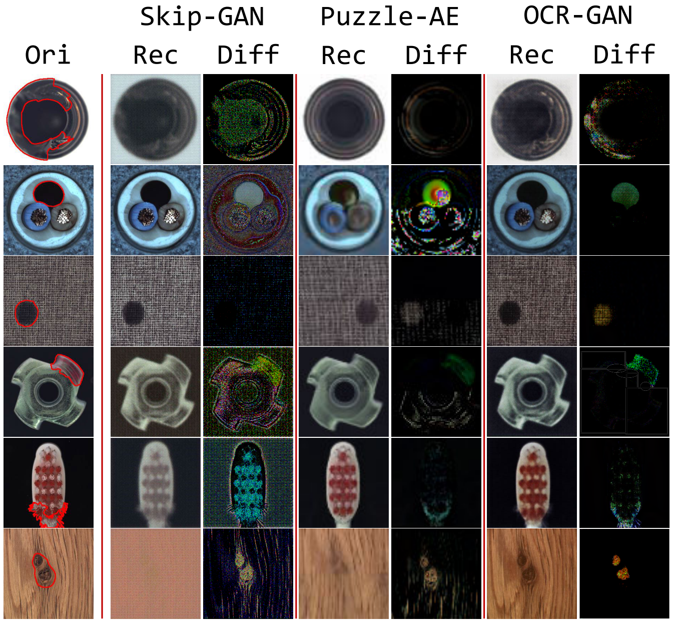


Fig. 12. **Reconstruction results of three reconstruction-based methods.** Ori: Original images with anomaly segmentation ground truth. Rec: Reconstructed images. Diff: Differences between original and reconstructed images.

on the model performance. As shown in Fig. 11, when we reduce the number of feature channels, the model performs better. *Our model performs best when the number of channels is set to 4, achieving 98.7 AUC on the MVTec AD dataset. This means that our lightweight model can even perform better in anomaly detection task.* Since the lightweight model is not the focus of this study, we will further fully study the design of the lightweight anomaly detection model and why the lightweight model can bring certain performance improvement in the future work. For fair comparison, all of our experiments in this paper (except this one) use the standard model, *i.e.*, the number of feature channels equals 64.

D. Interpretability of OCR-GAN

1) Analysis of Histogram: We visualize the anomaly score histogram for each category to further prove the effectiveness of our OCR-GAN. As shown in Fig. 8, abnormal samples would get higher anomaly scores while normal samples get lower anomaly scores, and there is a clear distinction between normal and abnormal samples, meaning that our model can well distinguish abnormal samples from normal samples by the anomaly score. Fig. 9 shows that normal samples and abnormal samples can not be distinguished by anomaly score in the histogram of the baseline. We further assess our proposed FD and CS module and results similarly indicate that each module contributes to the model.

2) Visualization of Latent-Space Features: We map the latent-space features from the last convolution layer of the D for each test sample to a two-dimensional subspace. Fig. 10 shows that our proposed OCR-GAN yields promising separation between normal and abnormal samples in the latent space.

3) Reconstruction Results: The reconstruction ability of the generator has a significant influence on the performance of the GAN-based method in the anomaly detection task. As shown in Fig. 12, we visualize the reconstructed images and the difference images between the reconstructed and original images to explore the reconstruction ability of different methods. Reconstructed results indicate that our OCR-GAN has a better reconstruction ability for details, and the abnormal areas are more prominent in the difference images than other classical reconstruction-based methods.

V. Conclusion

This paper proposes a novel reconstruction-based OCR-GAN for anomaly detection from a perspective of frequency domain. Specifically, we propose FD module to decouple the input image into different frequencies and model the reconstruction process as a combination of parallel omni-frequency image restorations. To better perform frequency interaction among different encoders, we propose a tailored CS module to adaptively select different channels among multiple branches. Our approach achieves new SOTA results over current SOTA methods on both sensory AD and semantic AD tasks even without extra training data, meaning that the proposed OCR-GAN is robust and effective for practical applications.

In the future, we will further explore the design of the lightweight model for AD tasks, while building more difficult practical dataset and testing the corresponding effects of our and other methods, hoping to contribute to the development of this field.

Acknowledgment

The authors would like to thank all the reviewers and the editor for their excellent contributions.

References

[1] M.-E. Nilsback and A. Zisserman, “Automated flower classification over a large number of classes,” in *Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process.*, Dec. 2008, pp. 722–729.

[2] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.

[3] X. Wang and M. Mirmehdi, “Archive film defect detection and removal: An automatic restoration framework,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3757–3769, Aug. 2012.

[4] Z. Li et al., "Thoracic disease identification and localization with limited supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8290–8299.

[5] Y. Liu, C.-L. Li, and B. Poczos, "Classifier two sample test for video anomaly detections," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Newcastle, U.K., Sep. 2018, p. 71.

[6] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6479–6488.

[7] H. Lv, C. Zhou, Z. Cui, C. Xu, Y. Li, and J. Yang, "Localizing anomalies from weakly-labeled videos," *IEEE Trans. Image Process.*, vol. 30, pp. 4505–4515, 2021.

[8] E. Jardim, L. A. Thomaz, E. A. B. da Silva, and S. L. Netto, "Domain-transformable sparse representation for anomaly detection in moving-camera videos," *IEEE Trans. Image Process.*, vol. 29, pp. 1329–1343, 2020.

[9] D. Tabernik, S. Sela, J. Skvarč, and D. Skočaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, May 2019.

[10] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," M.S. thesis, Dept. Comput. Sci., Citeseer, Univ. Toronto, Toronto, ON, Canada, 2009.

[11] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," 2020, *arXiv:2005.02357*.

[12] J. Yi and S. Yoon, "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–16.

[13] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A patch distribution modeling framework for anomaly detection and localization," in *Proc. Int. Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2021, pp. 475–489.

[14] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9659–9669.

[15] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial Nets," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1577–1581.

[16] T. N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1273–1283.

[17] X. Xia, X. Pan, X. He, J. Zhang, N. Ding, and L. Ma, "Discriminative-generative representation learning for one-class anomaly detection," 2021, *arXiv:2107.12753*.

[18] B. Hu et al., "A lightweight spatial and temporal multi-feature fusion network for defect detection," *IEEE Trans. Image Process.*, vol. 30, pp. 472–486, 2020.

[19] V. Zavrtanik, M. Kristan, and D. Skocaj, "DRÆM—A discriminatively trained reconstruction embedding for surface anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8330–8339.

[20] X. Li, X. Jin, T. Yu, S. Sun, Y. Pang, Z. Zhang, and Z. Chen, "Learning omni-frequency region-adaptive representations for real image super-resolution," in *Proc. 34th (AAAI) Conf. Artif. Intell. (AAAI), 33rd Conf. Innov. Appl. Artif. Intell. (IAAI)*, 2021, vol. 35, no. 3, pp. 1975–1983.

[21] S. Akçay, A. Atapour-Abarghouei, and T. P. Breckon, "Skip-ganomaly: Skip connected and adversarially trained encoder–decoder anomaly detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2019, pp. 1–8.

[22] J. Andrews, T. Tanay, E. J. Morton, and L. D. Griffin, "Transfer representation-learning for anomaly detection," in *Proc. JMLR*, 2016, pp. 1–5.

[23] T. S. Nazare, R. F. de Mello, and M. A. Ponti, "Are pre-trained CNNs good feature extractors for anomaly detection in surveillance videos?" 2018, *arXiv:1811.08495*.

[24] L. Bergman and Y. Hoshen, "Classification-based anomaly detection for general data," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–12.

[25] O. Rippel, P. Mertens, and D. Merhof, "Modeling the distribution of normal data in pre-trained deep features for anomaly detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 6726–6733.

[26] J. Yang, Y. Shi, and Z. Qi, "DFR: Deep feature reconstruction for unsupervised anomaly segmentation," 2020, *arXiv:2012.07122*.

[27] Z. Zeng, B. Liu, J. Fu, and H. Chao, "Reference-based defect detection network," *IEEE Trans. Image Process.*, vol. 30, pp. 6637–6647, 2021.

[28] T. Zhang, A. Wiliem, and B. C. Lovell, "Region-based anomaly localisation in crowded scenes via trajectory analysis and path prediction," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2013, pp. 1–7.

[29] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4182–4191.

[30] Y. Chen, Y. Tian, G. Pang, and G. Carneiro, "Unsupervised anomaly detection with multi-scale interpolated Gaussian descriptors," 2021, *arXiv:2101.10043*.

[31] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 255–262.

[32] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, Apr. 1984.

[33] M. Turcotte, J. Moore, N. Heard, and A. McPhall, "Poisson factorization for peer-based anomaly detection," in *Proc. IEEE Conf. Intell. Secur. Informat. (ISI)*, Sep. 2016, pp. 208–210.

[34] L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[35] L. Bergman, N. Cohen, and Y. Hoshen, "Deep nearest neighbor anomaly detection," 2020, *arXiv:2002.10445*.

[36] Y. Chen, X. Sean Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. Int. Conf. Image Process.*, 2001, pp. 34–37.

[37] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Comput. Vis. Image Understand.*, vol. 172, pp. 88–97, Jul. 2018.

[38] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14298–14308.

[39] M. Rudolph, B. Wandt, and B. Rosenhahn, "Same same but DifferNet: Semi-supervised defect detection with normalizing flows," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1906–1915.

[40] F. Ju, Y. Sun, J. Gao, Y. Hu, and B. Yin, "Image outlier detection and feature extraction via L1-norm-based 2D probabilistic PCA," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4834–4846, Dec. 2015.

[41] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 413–422.

[42] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[43] J. Tack, S. Mo, J. Jeong, and J. Shin, "CSI: Novelty detection via contrastive learning on distributionally shifted instances," in *Proc. 34th Conf. Neural Inf. Process. Syst.*, 2020, pp. 11839–11852.

[44] Y. Xiao, H. Wang, W. Xu, and J. Zhou, "L1 norm based KPCA for novelty detection," *Pattern Recognit.*, vol. 46, no. 1, pp. 389–396, Jan. 2013.

[45] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "LREN: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 35, no. 5, pp. 4139–4146.

[46] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Represent.*, vol. 14, 2014.

[47] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[48] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, vol. 5, Prague, Czech Republic, Feb. 2019, pp. 1–8.

[49] D. Pathak, P. Krähenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[50] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3379–3388.

[51] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 622–637.

[52] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 481–490.

[53] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2893–2901.

[54] M. Zaigham Zaheer, J.-H. Lee, M. Astrid, and S.-I. Lee, "Old is gold: Redefining the adversarially learned one-class classifier training paradigm," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14171–14181.

[55] G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated gradient representations for anomaly detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 206–226.

[56] X. Han, X. Chen, and L.-P. Liu, "GAN ensemble for anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 5, pp. 4090–4097.

[57] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 510–519.

[58] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.

[59] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 146–157.

[60] M. Wieler and T. Hahn, "Weakly supervised learning for industrial optical inspection," in *Proc. DAGM Symp.*, vol. 6, 2007.

[61] M. Salehi, A. Eftekhar, N. Sadjadi, M. H. Rohban, and H. R. Rabiee, "Puzzle-AE: Novelty detection in images through solving puzzles," 2020, *arXiv:2008.12959*.

[62] J. Pirnay and K. Chai, "Inpainting transformer for anomaly detection," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2022, pp. 394–406.

[63] Z. Lin, H. Ye, B. Zhan, and X. Huang, "An efficient network for surface defect detection," *Appl. Sci.*, vol. 10, no. 17, p. 6085, Sep. 2020.

[64] J. Božic, D. Tabernik, and D. Skocaj, "End-to-end training of a two-stage neural network for defect detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5619–5626.

[65] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, May 2019, pp. 8026–8037.

[66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.

[67] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.

Yufei Liang received the B.S. and M.S. degrees in control science and engineering from Zhejiang University, Zhejiang, China, in 2020 and 2023, respectively. Her research interests include anomaly detection, computer vision, and deep learning.

Jiangning Zhang received the B.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2017, and the Ph.D. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2022. His research interests include artificial intelligence generated content, anomaly detection, and deep learning.

Shiwei Zhao received the M.S. degree from Zhejiang University, Hangzhou, China, in 2019. He is currently a Researcher with the NetEase Fuxi AI Laboratory, Hangzhou, China. His research interests include user profiling, anomaly detection, deep learning, and their application in games.

Runze Wu received the Ph.D. degree from the University of Science and Technology of China. He is currently a Senior Researcher and the Head of the User Profiling Research Group, NetEase FUXI AI Laboratory, China. He has published more than 50 papers in refereed journals and conference proceedings, such as IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Transactions on Multimedia (TMM), *ACM Transactions on Information Systems* (TOIS), *ACM Transactions on Knowledge Discovery from Data* (TKDD), ACM SIGKDD, ACM MM, IJCAI, AAAI, TheWebConf, and ACM CIKM. He has served regularly on the program committees of conferences, including ACM SIGKDD, AAAI, IJCAI, TheWebConf, ACM CIKM, and IEEE ICDM. His research interests include user profiling, anomaly detection, causal inference, combinatorial optimization, deep learning, and various data mining and artificial intelligence applications across online games. Please find more information at https://wu-runze.github.io/.

Yong Liu (Member, IEEE) received the B.S. degree in computer science and engineering and the Ph.D. degree in computer science from Zhejiang University, Zhejiang, China, in 2001 and 2007, respectively. He is currently a Professor with the Institute of Cyber-Systems and Control, Zhejiang University. His main research interests include robot perception and vision, deep learning, big data analysis, multisensor fusion, machine learning, computer vision, information fusion, and robotics.

Shuwen Pan received the B.S. and master's degrees in industrial and electrical automation from Zhejiang University, Zhejiang, China, in 1996 and 2001, respectively, and the Ph.D. degree in structural engineering and health monitoring from the University of California at Irvine, Irvine, CA, USA, in 2006. He is currently a Professor with the School of Information and Electrical Engineering, Hangzhou City University. His main research interests include system identification, computer vision, robotic control, and robotics.