Unsupervised Image Anomaly Detection and Segmentation Based on Pre-trained Feature Mapping

Qian Wan, Liang Gao, *Senior Member, IEEE*, Xinyu Li, *Member, IEEE*, and Long Wen

***Abstract*—Image anomaly detection and segmentation are important for the development of automatic product quality inspection in intelligent manufacturing. Because the normal data can be collected easily and abnormal ones are rarely existent, unsupervised methods based on reconstruction and embedding have been mainly studied for anomaly detection. But the detection performance and computing time requires to be further improved. This paper proposes a novel framework, named as Pre-trained Feature Mapping (PFM), for unsupervised image anomaly detection and segmentation. The proposed PFM maps the image from a pre-trained feature space to another one to detect the anomalies effectively. The bidirectional and multi-hierarchical bidirectional pre-trained feature mapping are further proposed and studied for improving the performance. The proposed framework achieves the better results on well-known MVTec AD dataset compared with state-of-the-art methods, with the area under the receiver operating characteristic curve of 97.5% for anomaly detection and of 97.3% for anomaly segmentation over all 15 categories. The proposed framework is also superior in term of the computing time. The extensive experiments on ablation studies are also conducted to show the effectiveness and efficient of the proposed framework.**

***Index Terms*—Anomaly Detection, Anomaly Segmentation, Pre-trained Feature Mapping.**

I. Introduction

IMAGE anomalies include the visual different or broken regions compared against the normal images. The detection of the anomalies are significant for the improvement of the automatic product quality inspection in any intelligent manufacturing system [1]–[4], can largely reduce human labor and improve the efficiency of the inspection.

The image anomaly detection task is to judge whether the image is normal or abnormal and the image anomaly segmentation task is to segment out the abnormal regions of the image in pixel-wise. The two tasks are long-standing problems in automatic industrial scenarios and are difficult to solve due to the unknowable distribution of the anomalies [5]. This is because the normal data are easy to obtain but the abnormal data are difficult to collect, and the human-labeled information about anomalies are almost non-existent in most scenarios [5], increasing the difficulties to detect the anomalies. More, the wide range, such as type and size and so on, of the anomalies further increases difficulty of the challenging. Fig.1 shows the various annoying anomalies of different industrial products on MVTec AD dataset [6], and their corresponding qualitative results detected by the proposed framework.

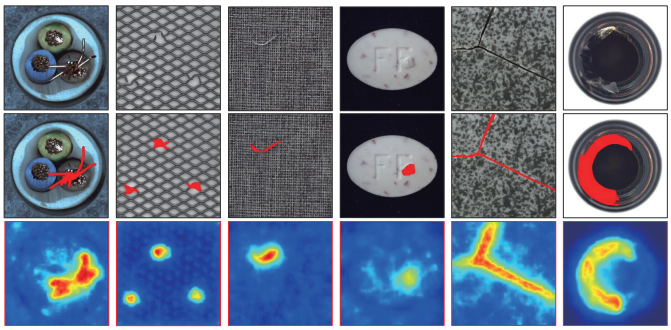
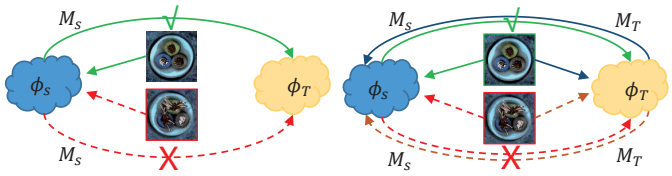


Fig. 1 Qualitative results of the proposed PFM on MVTec AD dataset [6]. Top Row: The testing image with anomalies (defects). Mid Row: The ground-truth of anomalies in red color. Bottom Row: The anomaly scoring maps are segmented in pixel-wise by the proposed PFM, and the anomalies contained in images are scored with higher values.

Recently, reconstruction [6]–[11] and embedding [12]–[16] based methods have been mainly studied for unsupervised image anomaly detection and segmentation. The reconstruction based methods apply the deep convolutional neural network (DCNN) to reconstruct the image and the difference between the original input and the reconstructed one can be computed to distinguish the anomalies. These types of methods are effective for the anomaly detection and segmentation on regular texture categories [8]. Instead of trying to fit the distribution of normal images by training the reconstruction procedure, the embedding based methods map the image to an embedded feature space to discriminate the anomalies. The Patch SVDD method [13] constructs the embedded feature space by self-training in a one-class classification manner and the Sub-Image Anomaly Detection with Deep Pyramid Correspondences (SPADE) method [15] directly adopts the pre-trained convolutional neural network as the embedded function. Both two effective embedding based methods detect the anomalies by retrieving the similar embedded features based on the KNN method, thus requiring a large amount of time computating and online memory storage. Although the Patch SVDD and SPADE method based on embedding have

This work was supported by National Natural Science Foundation of China under Grant No. 52188102 and U21B2029, and the Major Special Science and Technology Project of Hubei Province under Grant No. 2020AEA009. *(Corresponding author: Xinyu Li.)*

Q. Wan, X.Y. Li, and L. Gao are with the State Key Laboratory of Digital Manufacturing Equipment & Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China; L. Wen is with School of Mechanical Engineering & Electronic Information, China University of Geosciences (Wuhan). (e-mail: wanqian19@hust.edu.cn, lixinyu@mail.hust.edu.cn, gaoliang@mail.hust.edu.cn, wenlong@cug.edu.cn).



(a) (b)

Fig. 2 Detecting image anomaly by space mapping. (a) Detecting based on mapping image from one space to another one. (b) Detecting based on bidirectional mapping.

a superior in the detecting accuracy, the large computing time impedes them applied in the real-world industrial scenarios.

Both Patch SVDD and SPADE methods can be viewed as finding a discriminative space to distinguish the anomalies from normal images and can be denoted as following:

$$\begin{cases} A(\mathbf{x}) = 0, \mathbf{x} \in \phi \\ A(\mathbf{x}) = 1, \mathbf{x} \notin \phi \end{cases} \tag{1}$$

where denotes the image is normal and denotes the image is abnormal. While in this paper, based on embedding, the proposed Pre-trained Feature Mapping (PFM) framework detects the anomalies by mapping the image from one pre-trained feature space to another one. The main idea of the proposed PFM is shown in Fig.2. Fig.2a shows that the normal image can be freely mapped from feature space to feature space by the mapping function, while the abnormal one cannot be mapped. This can be denoted as following:

$$\begin{cases} A(\mathbf{x}) = 0, M_S(\phi_S(\mathbf{x})) \in \phi_T \\ A(\mathbf{x}) = 1, M_S(\phi_S(\mathbf{x})) \notin \phi_T \end{cases} \tag{2}$$

Fig.2b shows the mapping procedure can be in a bidirectional manner and denoted as following:

$$\begin{cases} A(\mathbf{x}) = 0, M_S(\phi_S(\mathbf{x})) \in \phi_T \;\&\&\; M_T(\phi_T(\mathbf{x})) \in \phi_S \\ A(\mathbf{x}) = 1, M_S(\phi_S(\mathbf{x})) \notin \phi_T \;||\; M_T(\phi_T(\mathbf{x})) \notin \phi_S \end{cases} \tag{3}$$

The proposed PFM framework improves the performance while also does not require time-consuming retrieval. This further boosts the application in real-word industrial scenarios.

The main contributions in this paper are listed as following:

- The novel pre-trained feature mapping framework is proposed for unsupervised image anomaly detection and segmentation.
- The extended bidirectional pre-trained feature mapping and multi-hierarchical bidirectional pre-trained feature mapping frameworks are also proposed and studied.
- The proposed framework achieves the improved performance on the anomaly detecting results on MVTec AD dataset compared with the state-of-the-art methods, with the area under the receiver operating characteristic curve (ROCAUC) of 97.5% for anomaly detection and of 97.3% for anomaly segmentation over all 15 categories. And it is also superior in term of the computing time.

In this paper, extensive experiments of ablation studies have also been conducted to show the effectiveness and efficient of the proposed PFM framework.

The remaining of this paper is organized as following: The related work about image anomaly detection and segmentation are stated in section II. The proposed PFM is introduced in section III. The extensive experiments are conducted and analyzed in section IV. The conclusion and future work are finally discussed in section V.

II. Related Work

The methods for image anomaly detection and segmentation can be mainly categorized into two types, which are reconstruction-based and embedding-based methods.

A. Reconstruction-based Methods

Reconstruction-based method have been widely studied for image anomaly detection [17], [18] and segmentation [6], [7]. The main idea behinds them is that to fit the distribution of the normal images by building the reconstruction network based on the all normal images. The differences between the input with the reconstructed one are scored for detection.

Deep Auto Encoder (AE) [19] is a powerful method to model the data in high dimensional space in unsupervised way. The AE has been studied for the anomaly segmentation [6], and the structural similarity [20] loss is applied to improve the vanilla AE for structural broken anomalies [7]. The drawback of the vanilla AE is that it can also generalize well for anomaly reconstruction and then miss the anomalies. The MemAE method [18] proposes a memory module to mostly focus on normal data reconstruction to mitigate the drawback of the vanilla AE. The retrieved similar memorized features for encoded features are decoded for the reconstruction, rather than directly apply the encoded feature for the reconstruction. Recently, the MemAE is combined with the perceptual loss [21] for anomaly segmentation [22]. Considering that the local anomalous regions can deteriorate the whole reconstruction, the iterative energy-based projection with gradient descent is proposed to improve the reconstruction quality [8]. The testing image of the AE is iteratively updated and the loss of high-frequency information caused by the AE bottleneck can be bypassed, leading to well reconstruction though among the heavily anomalous regions. Convolutional adversarial variational auto encoder with guided attention (CAVGA) method [9] localizes the anomaly with a convolutional latent variable to preserve the spatial information. They incorporate the Grad-CAM [23] into the variational AE with the expansion loss to make the model focus on the normal regions during reconstruction. Then an anomalous attention map can be obtained from the convolutional latent variable by applying the Grad-CAM to segment out the anomalies. The P-Net framework [10] is proposed to leverage the relation between the image texture and structure for anomaly detection and segmentation, to decrease the influence of the anomalous regions during the reconstruction. Artificial broken patches, simulated as anomalous regions, are used in the training stage to improve the anomaly suppression ability and performs well on anomaly localization [11]. The AE is trained by randomly removing partial local regions of the training image and reconstruct the image from the partial inpainting patches, to artificially address the drawback that the anomaly pixels are not visible to the AE

during training, and further improves the reconstruction quality when testing an anomalous image.

The reconstruction-based methods are easy to understand and effective for the regular texture categories, and most of them try to mitigate the drawback that the anomalous regions can deteriorate the whole reconstruction but cannot be avoided [8], [11].

B. Embedding-based Methods

Embedding-based methods benefit from the well discriminative embedding space, where the normal images can be embedded into a very compressed space. The embedding of the abnormal images are sparse and far away from the normal clusters in the embedding space [24].

Convolutional neural network has been heavily studied for the supervised image classification, many effective DCNN frameworks have been proposed, such as VGG [25] and ResNet [26]. The pre-trained DCNN on large image dataset can be directly transferred in other tasks, such as clustering [27]. Recently, the pre-trained ResNet on ImageNet dataset has been applied into anomaly detection [28], and shows that the pre-trained DCNN can be an effective and discriminative embedding function. In the CNN_Dict method [12] applies the pre-trained ResNet18 to extract the features of the images patch-by-patch, and all the normal features are decomposed by PCA method and clustered into several centers by k-means clustering method. Then the features of the testing image are also decomposed in the same way and the distances between the centers are averaged as the criterion as the score for anomaly detection and segmentation. The Student-Teacher (ST) framework [14] is proposed to detect anomalies by computing the differences between the output of the student network with that of the teacher network. The teacher network can be a pre-trained DCNN or trained by self-supervised learning and the student networks are trained to regress the output of the teacher network. The ST finally performs the anomaly detection and segmentation patch-by-patch. The SPADE method [15] utilizes the pre-trained Wide-ResNet [29] to extract the features of the image in multi-hierarchies and retrieves the several most similar features for testing features of the testing image from the normal features library. And the average distances between the testing feature and retrieved features are viewed as the score for anomaly detection and segmentation. The SPADE shows an effective performance on the two tasks but requires twice retrievals which consume much more computation time and online memory storage as the amount of the normal images increases. Differing from the SPADE, the Patch SVDD method [13] constructs the embedding function by self-supervised learning. The Patch SVDD also requires much computation time for retrieval on anomaly detection and segmentation. The embedding based methods, such as Patch SVDD [13] and SPADE [15], show the effective performance on anomaly detection and segmentation, but still require much more online memory storage and computing time as the amount of the normal images increases.

The proposed PFM framework based on embedding performs anomaly detection and segmentation by mapping the pre-trained feature from one space to another space as Fig.2 shows, achieving the better performance and further reducing the computing time.

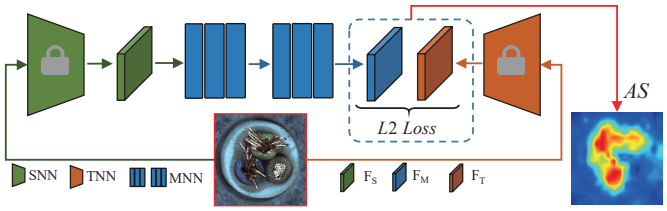


Fig. 3 The framework of Pre-trained Feature Mapping for image anomaly detection and segmentation.

III. The Proposed PFM Framework

The proposed PFM framework is elaborately introduced in this section. Firstly, the basic PFM is introduced. Then, considering the bidirectional mapping, the bidirectional PFM is introduced. And the multi-hierarchical bidirectional PFM is finally introduced, to take full use of the multi-hierarchical information to improve the performance.

A. Pre-trained Fature Mapping

The main idea of the proposed PFM is to map image from one pre-trained space to another one. The two spaces are denoted as source space and target space. As the Eq.2 denotes, the normal images can be freely mapped from the source space ϕ_S to target space ϕ_T, while the abnormal ones cannot be mapped. Fig.3 shows the framework of the proposed PFM.

The pre-trained DCNN model has shown the effective embedding ability for clustering [27] and anomaly detection [28]. In this paper, different from the methods like SPADE, it is proposed to construct a mapping neural network to map the pre-trained features from one pre-trained space to another pre-trained space only for normal images during the training stage in our proposed PFM framework. This means that the optimized mapping neural network can bridge up the gap between the two different pre-trained DCNN models only for normal images. To guarantee the discriminative fidelity of the PFM framework, the mapping neural network is optimized based on the normal features of normal training images by the mean squared loss during the training stage. Therefore, the gap for normal images between the two pre-trained feature spaces constructed by the two pre-trained DCNN models can be bridged up, while the anomalous image cannot be mapped from the space to another one by the optimized mapping neural network, so that the anomalies can cause a much larger mean square errors and be detected out.

The image is embedded into the source feature space by the pre-trained source neural network $\mathrm{SNN}(\cdot;\Theta_S)$ and embedded into the target feature space by the pre-trained target neural network $\mathrm{TNN}(\cdot;\Theta_T)$, and the corresponding embedding features are denoted as following:

$$\begin{aligned} F_S &= \mathrm{SNN}(\mathbf{x};\Theta_S) \\ F_T &= \mathrm{TNN}(\mathbf{x};\Theta_T) \end{aligned} \tag{4}$$

where $F_S \in \mathbb{R}^{w\times h\times c_S}$ denotes the source embedding feature map of the input image and $F_T \in \mathbb{R}^{w\times h\times c_T}$ denotes the target

one. For set of all normal training images $\cup \mathbf{x}$, two embedding feature map set can be constructed as source and target set, denoted as $\mathcal{F}_S = \cup F_S$ and $\mathcal{F}_T = \cup F_T$. Due to the well pre-trained DCNN model, the images can be embedded into the compacted and discriminative space. The issue that the labeled information is non-existent can be indirectly solved by the one-to-to mapping between $\mathcal{F}_S$ and $\mathcal{F}_T$. In this paper, the feature mapping function is implemented by the convolutional neural network, termed as $\mathrm{MNN}(\cdot;\Theta_M)$. The concrete structure of the MNN is introduced in section IV. And the mapping can be denoted as following:

$$\mathcal{F}_S \xrightarrow{\mathrm{MNN}} \mathcal{F}_T \tag{5}$$

for each normal training image, a mapping output feature map F_{M_S} can be computed as following:

$$F_{M_S} = \mathrm{MNN}(\hat{F}_S;\Theta_{M_S}) \tag{6}$$

where $\hat{F}_S(i,j) = \frac{F_S(i,j)}{\|F_S(i,j)\|_2}$, $\|\cdot\|_2$ denotes the spatial position in the feature map, denotes as the L_2 *norm*. The Θ_{M_S} can be optimized by the gradient descent method on normal training images, and the mapping loss object function is denoted as following:

$$Loss_{PFM} = \frac{1}{w \times h \times c_T} \left\| \hat{F}_{M_S} - \hat{F}_T \right\|_2^2 \tag{7}$$

where $\hat{F}_{M_S}(i,j) = \frac{F_{M_S}(i,j)}{\|F_{M_S}(i,j)\|_2}, \hat{F}_T(i,j) = \frac{F_T(i,j)}{\|F_T(i,j)\|_2}$. It is because that the MNN is only optimized on normal training images, then only normal images can be freely mapped from the source feature space to the target feature space but the mapping of abnormal images is blocked.

For a testing image, the anomaly scoring map $AS \in \mathbb{R}^{w \times h}$ can be computed by the mapping as following:

$$AS(i,j) = \left\| \hat{F}_{M_S}(i,j) - \hat{F}_T(i,j) \right\|_2^2 \tag{8}$$

For anomaly detection, the anomaly score can be computed for the testing image as a result as following:

$$S = \max(AS) \tag{9}$$

The maximum value in the anomaly segmentation map is adopted as the evaluated score for anomaly detection, and it is reasonable that the anomalous regions in the testing image always get much higher values in the scoring map. The single direction of the mapping is only used and the another reversed direction is ignored in PFM framework, this reduces the advantage of the utility of both directions. Thus, the Bidirectional PFM is proposed to improve the performance.

B. Bidirectional Pre-trained Feature Mapping

The number of the parameters of the mapping neural network in the PFM framework is much less than the one of the source neural network or the target neural network and the image is embedded into the feature space only once by the pre-trained source and target neural network. Thus, adding a reversed mapping neural network to the PFM framework would not carry any much more time consuming or online storage during the inference stage. So, the proposed Bidirectional Pre-trained Feature Mapping framework is proposed as shown in Fig.2. It is proposed to construct the two directional mapping neural networks with a part of shared parameters, and the whole parameters of the mapping networks are optimized by the both directional mean square losses based on the normal training images during the training stage. The B-PFM can be viewed as the bidirectional mapping ensemble of two PFM frameworks while brings less additional time and storage consuming, but the performance for image anomaly detection and segmentation is improved. In the PFM, only the one direction is used and the reversed one is dropped. While in the extended B-PFM, both the two directions are fully utilized for improving the anomaly detection and segmentation.

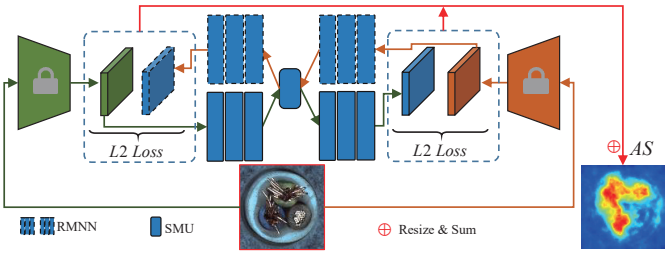


Fig. 4 The framework of the extended Bidirectional Pre-trained Feature Mapping for image anomaly detection and segmentation.

The reversed mapping is also conducted by a convolutional neural network, termed as $\mathrm{RMNN}(\cdot;\Theta_{M_T})$ and the bidirectional mapping is denoted as following:

$$\mathcal{F}_S \underset{\mathrm{RMNN}}{\overset{\mathrm{MNN}}{\Longleftrightarrow}} \mathcal{F}_T \tag{10}$$

The RMNN and MNN share a part of parameters, which is called Shared Mapping Unit (SMU). This design can decrease the amount of the parameters and it hopes that there should be common shared latent feature descriptors in both two directional mappings. Fig.4 shows the framework of the B-PFM. For each normal training image, the reversed mapping output feature map can be computed as following:

$$F_{M_T} = \mathrm{RMNN}(\hat{F}_T;\Theta_{M_T}) \tag{11}$$

Then the Θ_{M_S} and Θ_{M_T} can be optimized simultaneously on normal training images, and the mapping loss object function is denoted as following:

$$\begin{aligned} &Loss_{B-PFM} \\ &= \frac{1}{w \times h}\left(\frac{1}{c_T}\left\| \hat{F}_{M_S} - \hat{F}_T \right\|_2^2 + \frac{1}{c_S}\left\| \hat{F}_{M_T} - \hat{F}_S \right\|_2^2\right) \end{aligned} \tag{12}$$

Where $\hat{F}_{M_T}(i,j) = \frac{F_{M_T}(i,j)}{\|F_{M_T}(i,j)\|_2}$. Then the anomaly scoring map AS can be computed by the mapping as following:

$$\begin{aligned} &AS(i,j) \\ &= \frac{\left\| \hat{F}_{M_S}(i,j) - \hat{F}_T(i,j) \right\|_2^2 + \left\| \hat{F}_{M_T}(i,j) - \hat{F}_S(i,j) \right\|_2^2}{2} \end{aligned} \tag{13}$$

In the bidirectional mapping, both two directions can be used and improves the performance on anomaly detection and segmentation, but ignores the hierarchical information. The multi-hierarchical features can be applied to reduce the influence of the down-sampled resolution of deep feature map. Based on the B-PFM, the multi-hierarchical features are fused in the framework to further improve the performance.

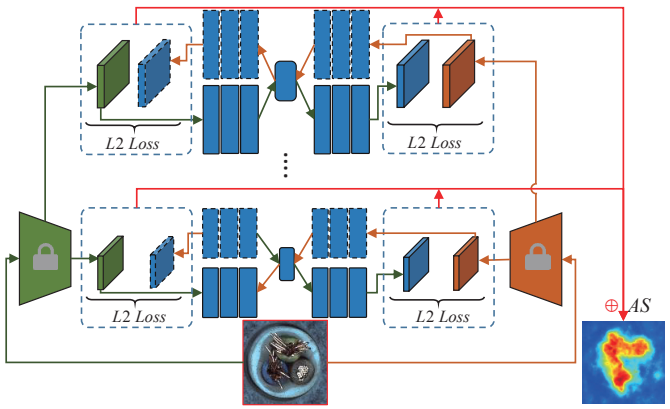


Fig. 5 The framework of the extended Multi-Hierarchical Bidirectional pre-trained Feature Mapping for image anomaly detection and segmentation.

C. Multi-Hierarchical Bidirectional Pre-trained Feature Mapping

The anomalies in images range over size and shape which may not be detected out in only one hierarchy. In the PFM and B-PFM, the hierarchical information is not considered and this can lead to miss anomalies. Multiple hierarchical features are widely used in many computer vision tasks such as object detection [30] and semantic segmentation [31], which have been approved effectively for multi-scale object detection or segmentation due to the different receptive fields of the different multi-hierarchical features. Thus, Multi-hierarchical Bidirectional Pre-trained Feature Mapping (MB-PFM) is further proposed to detect anomalies in different hierarchies. Fig.5 shows the framework of the MB-PFM. This further improves the performance for image anomalies detection and segmentation because the mappings in the different hierarchies have different receptive fields and can detect anomalies in multi-scale. The multi-hierarchical bidirectional mapping is proposed to bridge up the gaps between the pre-trained DCNN models in multi-hierarchy. The parameters of bidirectional mapping neural network in every hierarchy are optimized by the mean square losses based only on normal features. Thus, the anomalies can lead to larger mean square errors in every hierarchy during the testing stage.

The image is respectively embedded into the multi-hierarchies by the different hierarchical SNN and the TNN. For the k hierarchy, the corresponding feature map is denoted as F_S^k and F_T^k. The B-PFM can be constructed in each hierarchy and optimized on normal training images. The anomaly scoring map of the hierarchy can be computed by the mapping as following:

$$AS^k(i,j) = \frac{\left\|\hat{F}_{M_S}^k(i,j) - \hat{F}_T^k(i,j)\right\|_2^2 + \left\|\hat{F}_{M_T}^k(i,j) - \hat{F}_S^k(i,j)\right\|_2^2}{2} \quad (14)$$

The all hierarchical anomaly segmentation maps are resized to the same size of the testing image by bilinear interpolation, then averaged as the final scoring map as following:

$$AS = \frac{\sum_{k=1}^{K} \text{resize}(AS^k)}{K} \quad (15)$$

where K denotes the total number of the hierarchies.

The multi-hierarchical features are introduced into the proposed framework, then the multi-hierarchical discriminative information can be fully utilized for anomaly detection since that the shallow hierarchy can mitigate the drawback that the resolution decreases as the hierarchy goes to deeper, further improving the performance for different scale anomalies.

IV. Experiments

In this section, the proposed PFM framework is compared with the state-of-the-art methods and the ablation studies have been experimentally conducted and discussed.

A. Datasets and Implementation Details

MVTec AD Dataset was published by MVTec Software GmbH [6]. The MVTec AD dataset has been widely studied for unsupervised image anomaly detection and segmentation, which consists of 15 different categories. 3629 normal images are provided for training. 467 normal images and 1258 abnormal images are given for testing. The anomalies in the testing images are variant in scale, shape, color and etc., further increasing the difficulties. All the testing images are provided with pixel-level labeled ground-truth.

Implementation Details. The parameters of the SNN and TNN are frozen during training. The ResNet34 and ResNet50 network pre-trained on ImageNet dataset are respectively adopted the SNN and the TNN. The concrete structure of the MNN is reported in TABLE I and the one of RMNN is reversely symmetric with the MNN. In the B-PFM and MB-PFM, Layer3 is designed as the SMU. The MNN and RMNN are randomly initialized and optimized by the Adam algorithm with a learning rate of 0.0003, batch size of 8, and weight decay of 0.00001. The number of the training epochs is set as 200. All the images are resized to 256×256 and normalized by the standard deviation and mean value of the ImageNet dataset.

The number of total hierarchies is set as 3 in this paper. The three hierarchical features are respectively extracted at the conv2_x, conv3_x and conv4_x three layers if the SNN and TNN are the ResNet. Else if the SNN and TNN are the VGG, the three hierarchical features are respectively extracted at last three downsampling layers. The setting of the parameter c of different hierarchies are reported in TABLE II in details.

Evaluation Criterion. This paper follows the common used the area under the receiver operating characteristic curve (ROCAUC) for detection and segmentation [6]. Because ROCAUC criterion is in favor of the large anomalous regions, the normalized area under the per-region overlap curve (PROAUC) when the false positive rate is lower than 0.3 is also adopted as the segmentation evaluation criterion [14]. For ROCAUC or PROAUC criterion, the higher value means the model is better. For image-level detection, the true positive rate is defined as the percentage of images that were correctly classified as anomalous and the false positive is the one that were wrongly classified as anomalous. As for pixel-level segmentation, the true positive rate is defined as the percentage of pixels that were correctly classified as anomalous across the evaluated category, and the false positive rate is the one

TABLE I: The concrete structure of MNN.

Name		Kernel Size[1]
Layer1	ConvBlock[2]	$1\times1\times c_S\times(c_S/2+c)$
Layer2		$1\times1\times(c_S/2+c)\times2c$
Layer3		$1\times1\times2c_S\times c$
Layer4		$1\times1\times c\times2c$
Layer5		$1\times1\times2c\times(c_T/2+c)$
Layer6		$1\times1\times(c_T/2+c)\times c_T$

[1] The parameter of the latent dimension, value setting in term of the SNN and TNN.
[2] ConvBlock = ReLU(BatchNorm(Conv())), where ReLU is the nonlinear activation unit, BatchNorm is the normalization unit, and Conv is the convolutional unit.

TABLE II: The concrete structure of MNN.

M[1]	Var	R18-34[2]	R18-50	R34-50	V16-19
1	c	50	200		200
2		100	400		400
3		200	900		400

[1] M denotes the hierarchy number.
[2] R denotes the ResNet framework, and V denotes the VGG framework.

that were wrongly classified as anomalous. For comparing the computing time during the inference stage, the common used frame per second (fps) is adopted as the criterion. Higher fps means lower computing time and higher efficiency.

B. Comparisons with the State-of-the-art Methods

The proposed MB-PFM is compared with the long-standing baseline methods AS_SSIM [7], VAE [8], CNN_Dict [12], AnoGan [32], and the recently proposed state-of-the-art methods, ST [14], CAVGA [9], Patch SVDD [13], SPADE [15], MKD [16].

1) Image Level Anomaly Detection: As shown in TABLE III, the proposed MB-PFM is superior than the state-of-the-art (SOTA) methods on most categories of MVTec AD dataset, and achieves the better average result overall categories on image-level anomaly detection. Especially for the Carpet, Leather, Bottle, Hazelnut and Metal nut, the MB-PFM achieves the 100% ROCAUC for anomaly detection, outperforming the SOTA methods a lot. This is because that the anomaly scores of the anomalous regions are much higher than the normal regions by the multi-hierarchical bidirectional mapping, then the anomaly can be much easily detected.

2) Pixel Level Anomaly Segmentation: As shown in TABLE IV, for the comparison of the pixel-level segmentation on MVTec AD dataset, the MB-PFM method also achieves the better result compared with the SOTA methods over all categories in term of the ROCAUC evaluation criterion. This shows that the proposed MB-PFM can detect anomalies well in pixel-wise. Due to that the ROCAUC is in favor of the large anomalous regions [14], the MB-PFM is also evaluated under the PROAUC criterion. As shown in TABLE V, the MB-PFM still outperforms the SOTA methods over all categories on MVTec AD dataset for the anomaly segmentation. This means that the MB-PFM detects the small anomalies well. As Fig.6 shows, the qualitative visualization comparisons with the two SOTA method SPADE [15] and Patch SVDD [13], and the long-standing baseline AE_SSIM [7] method. From the visualizations of anomaly scoring maps, the MB-PFM focuses much more on anomalous regions. When the false positive is 0.5% over the same evaluated category, the segmented results

TABLE III: The image level anomaly detection results comparing with the SOTA methods on MVTec AD dataset (ROCAUC/%). Best in Bold.

Category	AE_SSIM [7]	CAVGA [9]	AnoGan [32]	VAE [8]	Patch SVDD [13]	MKD [16]	SPADE [15]	MB-PFM
Carpet	67.0	73.0	49.0	67.0	92.9	79.3	92.8	**100.0**
Grid	69.0	75.0	51.0	83.0	94.6	78.1	47.3	**98.0**
Leather	46.0	71.0	52.0	71.0	90.9	95.1	95.4	**100.0**
Tile	52.0	70.0	51.0	81.0	97.8	91.6	96.5	**99.6**
Wood	83.0	85.0	68.0	89.0	96.5	94.3	95.8	**99.5**
Bottle	88.0	89.0	69.0	86.0	98.6	99.4	97.2	**100.0**
Cable	61.0	63.0	53.0	56.0	90.3	89.2	84.8	**98.8**
Hazelnut	54.0	84.0	50.0	74.0	92.0	98.4	88.1	**100.0**
Metal nut	54.0	67.0	50.0	78.0	94.0	73.6	71.0	**100.0**
Pill	60.0	88.0	68.0	80.0	86.1	82.7	80.1	**96.5**
Screw	51.0	77.0	35.0	71.0	81.3	83.3	66.7	**91.8**
Toothbrush	74.0	91.0	57.0	89.0	**100.0**	92.2	88.9	88.6
Transistor	52.0	73.0	67.0	70.0	91.5	85.6	90.3	**97.8**
Zipper	80.0	87.0	59.0	67.0	97.9	93.2	96.6	**97.4**
Average	63.0	78.0	55.0	77.0	92.1	87.7	85.5	**97.5**

TABLE IV: The pixel level anomaly segmentation results comparing with the SOTA methods on MVTec AD dataset (ROCAUC/%). Best in Bold.

Category	AE_SSIM [7]	CAVGA [9]	AnoGan [32]	VAE [8]	Patch SVDD [13]	MKD [16]	SPADE [15]	MB-PFM
Carpet	87.0	-	54.0	73.5	92.6	95.6	97.5	**99.2**
Grid	94.0	-	58.0	96.1	96.2	91.8	93.7	**98.8**
Leather	78.0	-	64.0	92.5	97.4	98.1	97.6	**99.4**
Tile	59.0	-	50.0	65.4	91.4	82.8	87.4	**96.2**
Wood	73.0	-	62.0	83.8	90.8	84.8	88.5	**95.6**
Bottle	93.0	-	86.0	92.2	98.1	96.3	**98.4**	**98.4**
Cable	82.0	-	78.0	91.0	96.8	82.4	**97.2**	96.7
Capsule	94.0	-	84.0	91.7	95.8	95.9	**99.0**	98.3
Hazelnut	97.0	-	87.0	97.6	97.5	94.6	**99.1**	**99.1**
Metal nut	89.0	-	76.0	90.7	98.0	86.4	**98.1**	97.2
Pill	91.0	-	87.0	93.0	95.1	89.6	96.5	**97.2**
Screw	96.0	-	80.0	94.5	95.7	96.0	**98.9**	98.7
Toothbrush	92.0	-	90.0	98.5	98.1	96.1	97.9	**98.6**
Transistor	90.0	-	80.0	91.9	**97.0**	76.5	94.1	87.8
Zipper	88.0	-	78.0	86.9	95.1	93.9	96.5	**98.2**
Average	87.0	89.0	74.3	89.3	95.7	90.7	96.5	**97.3**

TABLE V: The pixel level anomaly segmentation results comparing with the SOTA methods on MVTec AD dataset (PROAUC/%). Best in Bold.

Category	AE_SSIM [7]	CNN_Dict [12]	AnoGan [32]	VAE [8]	ST [6]	SPADE [15]	MB-PFM
Carpet	64.7	46.9	20.4	50.1	87.9	94.7	**96.9**
Grid	84.9	18.3	22.6	22.4	95.2	86.7	**96.0**
Leather	56.1	64.1	37.8	63.5	94.5	97.2	**98.8**
Tile	17.5	79.7	17.7	87.0	**94.6**	75.9	88.7
Wood	60.5	62.1	38.6	62.8	91.1	87.4	**92.6**
Bottle	83.4	74.2	62.0	89.7	93.1	**95.5**	95.4
Cable	47.8	55.8	38.3	65.4	81.8	90.9	**94.2**
Capsule	86.0	30.6	30.6	52.6	**96.8**	93.7	91.7
Hazelnut	91.6	84.4	69.8	87.8	96.5	95.4	**96.7**
Metal nut	60.3	35.8	32.0	57.6	94.2	94.4	**94.6**
Pill	83.0	46.0	77.6	76.9	96.1	94.6	**96.1**
Screw	88.7	27.7	46.6	55.9	94.2	**96.0**	93.4
Toothbrush	78.4	15.1	74.9	69.3	93.3	**93.5**	90.7
Transistor	72.5	62.8	54.9	62.6	66.6	**87.4**	74.9
Zipper	66.5	70.3	46.7	54.9	95.1	92.6	**94.8**
Average	69.4	51.5	44.3	63.9	91.4	91.7	**93.0**

of the MB-PFM are more accurate than the three methods under the corresponding scoring threshold.

3) Processing efficiency: In this part, the proposed framework is compared with the effective SPADE method in term of the time computing. The testing image size is $256\times256\times3$. And The hardware environment is the computer of Intel i5 @2.9GHz CPU, Nvidia GTX 1660 GPU and 16G RAM. The MB-PFM is also superior than the state-of-the-art SPADE [15] method in term of the computing time. As shown in TABLE VI, the frame per second of the MB-PFM is 14.5, and approximately 242 times faster than the SPADE to obtain the anomaly scoring map.

Due to the complex structure of different convolutional neural network, the FLOPs (floating point operations per second) as common [26] is used to directly descript the time complexity in the actual computing during pre-trained

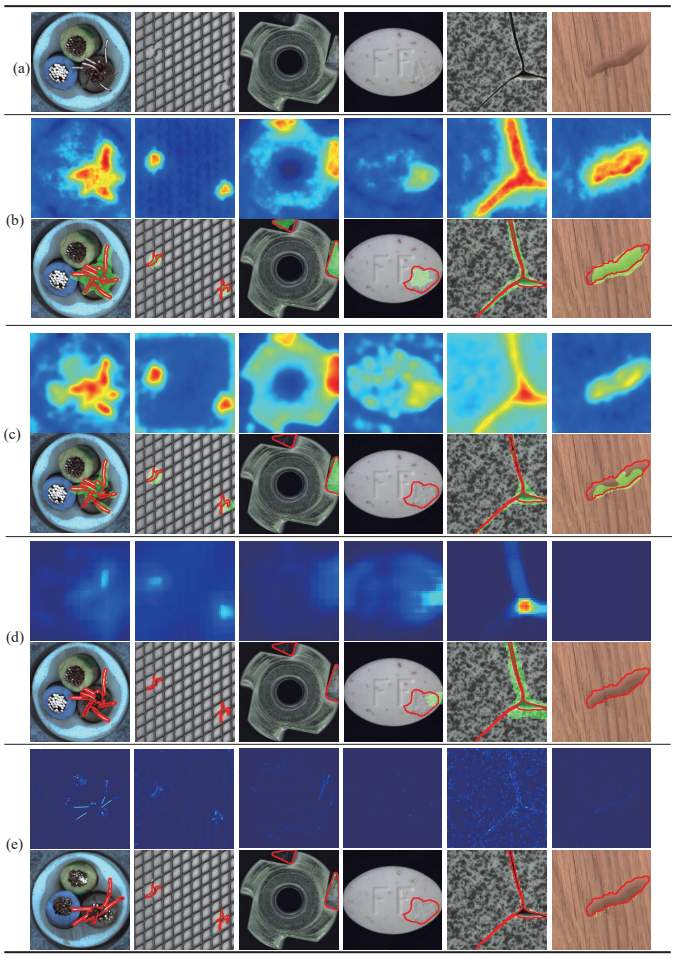


Fig. 6 Qualitative visualization comparisons on MVTec AD dataset [6]. (a) Testing anomalous images. (b)-(e): The corresponding anomaly scoring maps and segmentation maps of the MB-PFM, SPADE [15], Patch SVDD [13] and AE_SSIM [7]. The segmented result is at the 0.5% false positive rate on scoring map over the same categories. The ground-truth is in red color and the segmented result is in green color.

TABLE VI: The Comparisons for computing time during inference.(FPS)

	Patch SVDD [13]	SPADE [15]	MKD [16]	PFM	B-PFM	MB-PFM
fps	0.18	0.06	11.5	23.5	23.3	14.5

feature extraction in inference stage. For a testing image, the 14.9Gb FLOPs is used to make a feature extraction for SPADE with pre-trained wide_resnet50, while 22.9Gb FLOPs is used in our proposed MB-PFM with multi-hierarchical bidirectional feature mapping between pre-trained features extracted by resnet34 and resnet50. Though our proposed MB-PFM consumes more computing time in this aspect, this gap can be largely closed by the parallel computing on GPU device. However as for outputting an anomaly scoring map, the SPADE method requires an $O(log_2 n)$ time complexity at least due to the twice K-NN retrievals from an online pre-stored feature library of training images, but the proposed MB-PFM framework only requires $O(1)$ time complexity since the anomaly scoring map can be directly obtained after the distance calculating of pre-trained feature mapping. Therefore, the proposed MB-PFM method is also much superior in term of time consuming.

TABLE VII: The settings of the ablation studies and the corresponding results on MVTec AD dataset.

N	SNN	TNN	ROCAUC(/%)	PROAUC(/%)		B[2]	M[3]		
			Image	pixel			1	2	3
1	R18	R34	93.5	95.1	84.7				√
2	R18	R34	95.1	95.9	86.8	√			√
3	R18	R34	93.6	94.8	90.0	√		√	
4	R18	R34	82.8	90.8	82.3	√	√		
5	R18	R34	97.2	96.8	91.6	√		√	√
6	R18	R34	93.1	95.1	90.7	√	√	√	
7	R18	R34	96.7	97.0	91.1	√	√		√
8	R18	R34	97.4	97.0	92.5	√	√	√	√
9	R18[1]	R34[1]	58.1	50.8	18.0	√	√	√	√
10	R18	R50	97.0	97.2	**93.0**	√	√	√	√
11	R34	R50	**97.5**	**97.3**	**93.0**	√	√	√	√
12	V16	V19	94.2	94.7	86.9	√	√	√	√

[1] denotes that the parameters of the network are randomly initialized and then fixed.
[2] B denotes whether the bidirectional mapping is used.
[3] M denotes that which hierarchical feature or combination the hierarchical features is used.

C. Ablation Study

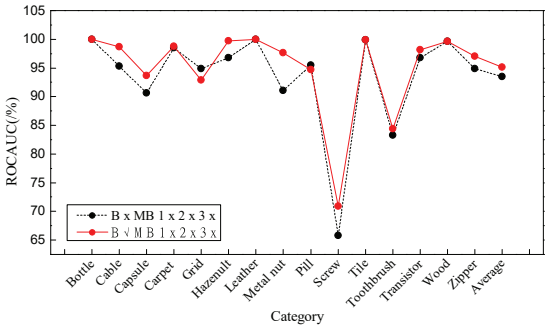
In this subsection, the extensive experiments of the ablation studies on MVTec AD dataset have been conducted and analyzed to show the superior of the proposed PFM framework. As reported in TABLE VII, the 12 instantiated cases have been experimentally studied in term of the bidirectional mapping, multi-hierarchical mappings and pre-trained DCNN models. The corresponding average results of the image-level anomaly detection and pixel-level segmentation over all 15 categories on MVTec AD dataset have been reported.

1) The Influence of Bidirectional Mapping: As the case 1 and 2 shown in TABLE VII, the influence of the bidirectional mapping has been studied. In this study, the SNN and TNN are respectively set as the pre-trained ResNet18 and ResNet34. And the third hierarchical feature is only used for mapping.

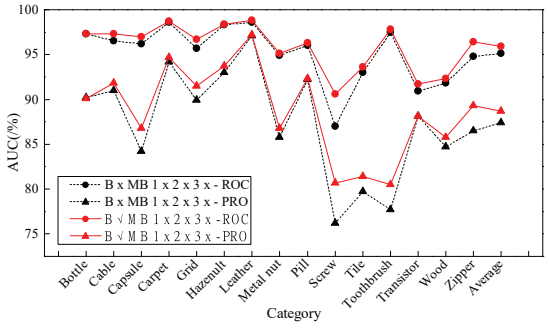
From the results of the case 1 and 2, the B-PFM is superior than the PFM in term of the image anomaly detection and segmentation. Also as Fig.7 shows, the results of the image-level anomaly detection and pixel-level segmentation for 15 categories are plotted. No matter for image-level or pixel-level, the B-PFM is always superior than the vanilla PFM. The B-PFM fully utilizes the both directional mapping information for anomaly detection and then improves the performance.

2) The Influence of Multi-Hierarchical Mappings: As the case 2 to 8 shown in TABLE VII, the influence of the multi-hierarchical mappings has been studied. In this study, the SNN and TNN are respectively set as the pre-trained ResNet18 and ResNet34. The three hierarchical features are studied for bidirectional mapping.

For the case 2, 3 and 4, the different hierarchy is independently studied. The performance of anomaly increases as the hierarchy goes to deeper. This is because the deeper hierarchical feature of the pre-trained models is more discriminative thus has a good distinguishable ability for anomaly. But for the pixel-level segmentation, the performance has some drop since that the resolution of the anomaly segmentation decreases as it goes to deeper, which can miss some anomalies. For the case 5, 6 and 7, the two combinations of the three hierarchies are studied. No matter for image-level or pixel-level, the performance has been improved compared with the case that only the single hierarchy is used. This shows that the combination

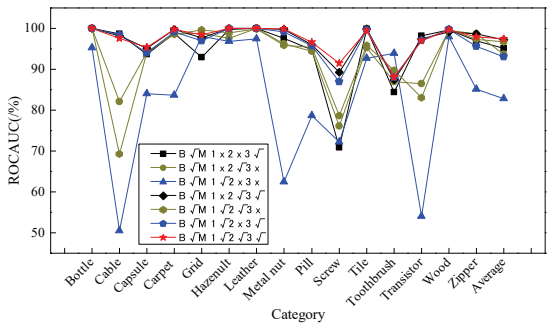


(a) Image Level

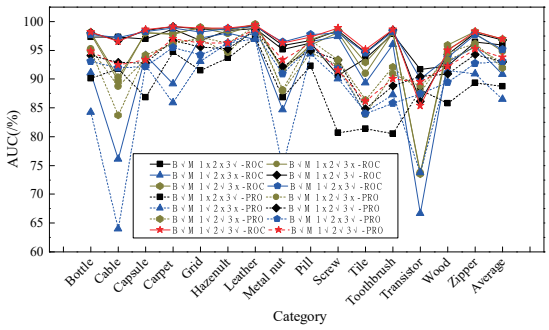


(b) Pixel Level

Fig. 7 The influence of the bidirectional mapping on image anomaly detection and segmentation.



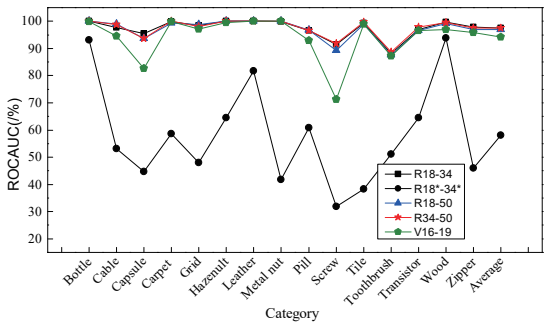
(a) Image Level



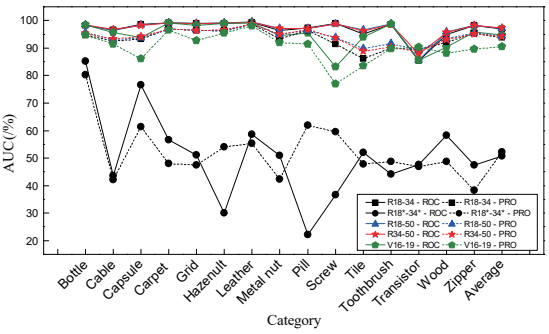
(b) Pixel Level

Fig. 8 The influence of the multi-hierarchical mappings on image anomaly detection and segmentation.

of the hierarchies can make an improvement for detection and segmentation. As for the case 8, both of the three hierarchies are utilized for the study. From the evaluation criterion, the performance has been generally improved overall categories. Also, as shown in Fig.8, the results of the detection and segmentation for 15 categories are plotted. From the figure, it



(a) Image Level



(b) Pixel Level

Fig. 9 The influence of the pre-trained SNN and TNN on image anomaly detection and segmentation.

can be seen that the case that both three hierarchies are used is generally superior than the others. The deeper hierarchy is more discriminative but the decreased resolution limits the performance on segmentation. Then, the combination of the shallow hierarchies can mitigate this drawback and improves performance well.

3) The Influence of Pre-Trained Networks: As the case 8 to 12 shown in TABLE VII, the influence of the pre-trained SNN and TNN has been studied. In this study, the bidirectional mapping of three hierarchical features are used.

For the case 8 and 9, the experiments that the random chosen parameters and pre-trained parameters on ImageNet of the SNN and TNN are studied. Both from the results of detection and segmentation, it shows that the pre-trained models are dramatically superior than the random ones. As shown in Fig.9, over all 15 categories, the case that the random parameters for SNN and TNN performs bad. The good embedding tool is important for the anomaly detection and the per-trained model embeds the image into the highly compressed and discriminative feature space. A pair of well pre-trained models is much significant for the proposed framework for detection and segmentation. The different pre-trained models have been studied in the case 8, 10, 11 and 12. The combination of VGG performs slightly not well than the ones of ResNet. The combination of the ResNet34 and ResNet50 generally performs better overall categories. The MNN and RMNN need to fit the gap between the SNN and TNN pre-trained models for the all normal training images. The larger gap makes the mapping procedure much harder. The mapping loss can be minimized during the training. But the mapping procedure is invalid for the unseen anomalies. Then, the larger gap mapping can make a higher difference on

unseen anomalies, thus improves the performance of anomaly detection and segmentation.

Over all the experimental studies on the bidirectional mapping, the multi-hierarchical mapping and the pre-trained model three aspects, it shows the superiority of the proposed PFM framework. The bidirectional mapping enhances the utilizing of the information in the two directions. The multi-hierarchical mapping mitigates the drawback that the decreased resolution of the feature map as the hierarchy goes to deeper. More, the pre-trained models on large dataset such as ImageNet are significant for the PFM framework, since that the image can be embedded into the compacted and discriminative feature space.

V. CONCLUSION

This paper proposed a novel framework named PFM for unsupervised image anomaly detection and segmentation. The B-PFM and MB-PFM methods are also proposed and studied to fully utilize the bidirectional mapping and multi-hierarchical information. The proposed framework achieves better performance on image anomaly detection and segmentation compared with the state-of-the-art methods on MVTec AD dataset. More, the proposed PFM framework is also superior in term of the computing time. The limitations of the proposed framework are in two aspects. First, the pre-trained models must be available while may not be adapted for some other dataset. Second, in some scenarios, the limited labeled information of anomalous data may exist, but the proposed framework only works in an unsupervised manner. In the future work, the adapted feature learning for the specified dataset and introducing the labeled information into the framework can be further studied to improve the performance on anomaly detection.

REFERENCES

[1] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7479–7488, 2020.

[2] Y. Zhang, Z. Y. Dong, W. Kong, and K. Meng, "A composite anomaly detection system for data-driven power plant condition monitoring," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4390–4402, 2020.

[3] A. Castellani, S. Schmitt, and S. Squartini, "Real-world anomaly detection by using digital twin systems and weakly supervised learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4733–4742, 2021.

[4] Y. Gao, L. Gao, and X. Li, "A generative adversarial network based deep learning method for low-quality defect image reconstruction and recognition," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3231–3240, 2020.

[5] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 756–795, 2021.

[6] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.

[7] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. 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*, 2019, pp. 372–380.

[8] D. Dehaene, O. Frigo, S. Combrexelle, and P. Eline, "Iterative energy-based projection on a normal data manifold for anomaly localization," in *International Conference on Learning Representations*, 2020.

[9] S. Venkataramanan, K.-C. Peng, R. V. Singh, and A. Mahalanobis, "Attention guided anomaly localization in images," in *European Conference on Computer Vision*, vol. 12362. Springer, 2020, pp. 485–503.

[10] K. Zhou, Y. Xiao, J. Yang, J. Cheng, W. Liu, W. Luo, Z. Gu, J. Liu, and S. Gao, "Encoding structure-texture relation with p-net for anomaly detection in retinal images," in *European Conference on Computer Vision*, vol. 12365. Springer, 2020, pp. 360–377.

[11] V. Zavrtanik, M. Kristan, and D. Skčaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, p. 107706, 2021.

[12] P. Napoletano, F. Piccoli, and R. Schettini, "Anomaly detection in nanofibrous materials by cnn-based self-similarity," *Sensors*, vol. 18, no. 1, p. 209, 2018.

[13] J. Yi and S. Yoon, "Patch svdd: Patch-level svdd for anomaly detection and segmentation," in *Proceedings of the Asian Conference on Computer Vision*, vol. 12627. Springer, 2020, pp. 375–390.

[14] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4183–4192.

[15] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," *CoRR*, vol. abs/2005.02357, 2020.

[16] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 902–14 912.

[17] Y. Xia, X. Cao, F. Wen, G. Hua, and J. Sun, "Learning discriminative reconstructions for unsupervised outlier removal," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1511–1519.

[18] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1705–1714.

[19] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.

[20] Z. Wang, A. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[22] D. S. Tan, Y.-C. Chen, T. P.-C. Chen, and W.-C. Chen, "Trustmae: A noise-resilient defect classification framework using memory-augmented auto-encoders with trust regions," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 276–285.

[23] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, pp. 336–359, 2019.

[24] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] J. Guérin and B. Boots, "Improving image clustering with multiple pretrained cnn feature extractors," in *Proceedings of the British Machine Vision Conference (BMCV)*, 2018.

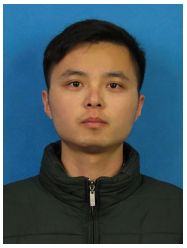
[28] L. Bergman, N. Cohen, and Y. Hoshen, "Deep nearest neighbor anomaly detection," *CoRR*, vol. abs/2002.10445, 2020.

[29] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proceedings of the British Machine Vision Conference (BMCV)*, 2016.

[30] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944.

[31] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6230–6239.

[32] 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 *Information Processing in Medical Imaging - 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings*, ser. Lecture Notes in Computer Science, vol. 10265. Springer, 2017, pp. 146–157.



Qian Wan was born in Jiangxi, China in 1996. He received the B.S. degree in 2018 and now is pursuing the Ph.D. degree in Mechanical and Electronic Engineering at Huazhong University of Science and Technology (HUST).

His current research interests include deep learning and its application in real industrial applications.

Liang Gao (M'08-S'20) received the B.Sc. degree in mechatronic engineering from Xidian University, Xi'an, China, in 1996, and the Ph.D. degree in mechatronic engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2002.

He is a Professor of the Department of Industrial and Manufacturing System Engineering, the Deputy Director of State Key Laboratory of Digital Manufacturing Equipment and Technology, and the Vice Dean of Research and Development Office, HUST. He was supported by the Program for New Century Excellent Talents in University in 2008 and the National Science Fund for Distinguished Young Scholars of China in 2018. His research interests include operations research and optimization, big data and machine learning etc. He has published more than 200 refereed papers indexed by SCIE, authored 7 monographs.

Dr. Gao currently serves as the Co-Editor-in-Chief for IET Collaborative Intelligent Manufacturing, an Associate Editor of Swarm and Evolutionary Computation, and the Journal of Industrial and Production Engineering.

Xinyu Li (M'19) received the Ph.D. degree in industrial engineering from Huazhong University of Science and Technology (HUST), China, in 2009.

He is a Professor of the Department of Industrial and Manufacturing Systems Engineering, State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, HUST. He has published more than 90 refereed papers. His research interests include intelligent algorithm, big data and machine learning etc.

Long Wen received the Ph.D. degree in industrial engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2014.

He is a Professor in the School of Mechanical Engineering and Electronic Information, China University of Geosciences. He has published more than 20 refereed papers, and his research interests include deep learning, automatic machine learning, fault diagnosis and intelligent algorithm, etc.