Logit Inducing With Abnormality Capturing for Semi-Supervised Image Anomaly Detection

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

***Abstract*— Image anomaly detection is a significant stage for visual quality inspection in intelligent manufacturing systems. According to the assumption that only normal images are available during the training stage, unsupervised methods have been studied recently for image anomaly detection. But anomalous images of small scale can be collected for training in many real-world industrial scenarios, and the unsupervised methods make no use of them to improve the detection accuracy. This leads to semi-supervised image anomaly detection with an unbalanced detection challenge. In this article, a logit inducing with abnormality capturing (LIAC) method is proposed to address semi-supervised image anomaly detection. First, a logit inducing loss (LIS) is proposed to train a classifier for dealing with unbalanced detection. Second, an abnormality capturing module (ACM) is proposed to address anomaly detection. With labeling only 40 anomalous images for training, the proposed LIAC method achieves a 98.8% F1-score on image anomaly detection of the printed circuit board, compared with the state-of-the-art (SOTA) methods. Moreover, the proposed LIAC method is experimentally compared with the SOTA methods on magnetic tile defect (MTD), retinal optical coherence tomography (ROCT), and electroluminescence images of photovoltaic modules (ELPV), three open-source datasets, respectively, which achieve F1-score of 85.2%, 96.8%, and 66.6% with given 40 anomalous images for training.**

***Index Terms*— Abnormality capturing, image anomaly detection, logit inducing, semi-supervised.**

I. Introduction

IMAGE anomaly detection is important for the development of visual quality inspection in intelligent manufacturing systems [1], [2]. The automatic method for image anomaly detection can largely reduce the human labor and help improve the processing efficiency of product [3], [4], [5]. Recently, unsupervised methods have been studied for image anomaly detection under the limitation that only normal images exist in the training stage [6], [7], [8], [9], [10], [11]. However, there are some anomalous images that can be collected in real-world industrial scenarios, and they can be used to improve the detection accuracy, but it is a pity that the unsupervised methods rarely make use of them. Given some anomalous images during training, it leads to a semi-supervised image anomaly detection problem with unbalanced detection.

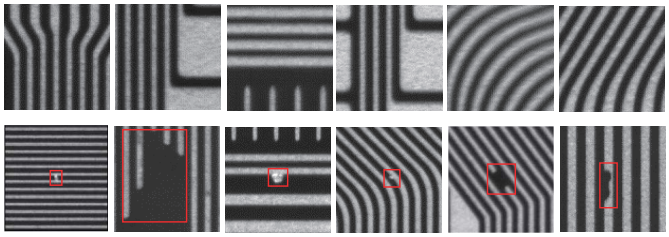


Fig. 1. Example images of the PCB dataset collected from a cooperative manufacturing company. The top rows are normal images. The bottom rows are anomalous images, in which the anomalies are circled with red boxes.

There are two challenges which are, respectively, unknown anomaly distribution and unbalanced detection to solve semi-supervised image anomaly detection. The anomalies in the anomalous images vary greatly in size, shape, position, etc. As shown in Fig. 1, all the images were collected from the production line of the printable circuit board (PCB) in a cooperative manufacturing company. The normal images can also change a lot in the texture structure, and the anomalies in the anomalous images widely range over size, shape, etc. The first, third, and fourth anomalous images in Fig. 1 are caused by over tin solder, which may lead to short-circuit accidents. The second, fifth, and sixth anomalous images are lack of tin solder, which could lead to open-circuit accidents. The unsupervised methods for image anomaly detection have been heavily studied [6], such as Gaussian clustering of pre-trained feature (GCPF) [7], student-teacher feature pyramid matching (STFPM) [12], pre-trained feature mapping (PFM) [11], conditional normalizing flows (CFlow) [13], and reverse distillation for anomaly detection (RD4AD) [14] methods. Although these methods perform well in capturing the normality distribution only with normal training images, they rarely make use of labeled anomalies to improve the detection accuracy. Some labeled anomalous images in the training set cannot provide enough information to capture the distribution of anomalies, but can provide clues to detect anomalies out [15]. These anomalies of small number cause an unbalanced problem in semi-supervised image anomaly detection which requires to be addressed well. For the general unbalanced image classification, many great methods based on loss adjustment have been studied [16], such as weighted cross entropy (WCE)-based [17], focal loss (FL)-based [18], sample-influence-based [19], balanced reference-distribution-based [20], and logit-adjustment-based [21]. These unbalanced methods require an improvement to perform well on semi-supervised image anomaly detection because the number of labeled training anomalies is small.

Manuscript received 14 July 2022; revised 23 August 2022; accepted 28 August 2022. Date of publication 12 September 2022; date of current version 23 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant U21B2029 and Grant 52188102 and in part by the Major Special Science and Technology Project of Hubei Province under Grant 2020AEA009. The Associate Editor coordinating the review process was Dr. Susanna Spinsante. *(Corresponding author: Xinyu Li.)*

The authors are with the State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: wanqian19@hust.edu.cn; gaoliang@mail.hust.edu.cn; lixinyu@mail.hust.edu.cn).

Digital Object Identifier 10.1109/TIM.2022.3205674

To address anomaly detection and unbalanced problem in semi-supervised image anomaly detection as stated above, a novel logit inducing with abnormality capturing (LIAC) method is proposed. In the proposed LICA method, a logit inducing loss (LIL) is proposed to train a classifier for unbalanced detection caused by some labeled training anomalous images. The LIL adjusts the classification logits with class frequencies and continually optimizes the parameters of a convolutional neural network with limited labeled anomalies. Second, a novel abnormality capturing module (ACM) is proposed to learn feature representation for capturing the anomaly. The ACM makes use of two backbone convolutional neural networks of the same structure to learn the feature representation by similarity and dissimilarity learning.

The main novelties and contributions in this article are listed as follows.

1) A novel LIAC method is proposed for semi-supervised image anomaly detection.
2) To address the unbalanced detection problem, a novel LIL is proposed to train a classifier to classify anomalies from normal images.
3) To address anomaly detection, a novel ACM is proposed to learn discriminative feature representation.
4) Extensive experiments on real application of PCB image anomaly detection and on open-source magnetic tile defect (MTD) [22], retinal optical coherence tomography (ROCT) [23], electroluminescence images of photovoltaic modules (ELPV) [24] datasets are conducted to show the improvement of the proposed method.

The remaining of this article is organized as follows. The related work is stated in Section II. The proposed LIAC method is introduced in Section III. The extensive experiments are conducted and analyzed with discussions in Section IV. The conclusion and future work are finally discussed in Section V.

II. Related Work

In this section, to help clearly state the proposed LIAC method for addressing semi-supervised image anomaly detection, related work about image anomaly detection and unbalanced image classification is discussed in detail.

A. Image Anomaly Detection

Image anomaly detection is challenging in real-world scenarios because anomalous images are limited and difficult to collect for training. Therefore, many kinds of research have been developed in an unsupervised manner based on only normal images to detect anomalies [6].

Deep metric learning based on a triplet network is proposed to detect anomalous images in a patchwise manner [25]. Patch-level support vector data description is proposed to detect anomalous images [26] by retrieving embeddings of the testing image from embeddings of negative samples by an optimized neural network which trained with support vector data description loss and self-supervised loss. The convolutional adversarial variational autoencoder with a convolutional latent expansion loss is also proposed to detect anomalous images [27]. Regression error between the student network and the teacher network in a patchwise way is used to detect anomalous images [28], in which the student network is trained based on the outputs of negative samples by the teacher network. Regression error between the student network and the teacher network in a patchwise way is used to detect anomalous images [28], in which the student network is trained based on the outputs of negative samples by the teacher network. Multivariate Gaussian distribution is applied to model pretrained features of negative samples to detect anomalous image [29]. The GCPF [7] method takes use of multiple multivariate Gaussian clusters to represent the prototypes of normal features and then applies the distance measurement to detect anomalies. The CFlow [13] method applies a conditional-flow-based model to fit the distribution of normal pretrained features, with probability away from distribution to score for anomalies. More, the semi-orthogonal embedding for anomaly detection (ORTHOAD) [30] method uses the semi-orthogonal embedding with random feature selection of pretrained features to detect anomalies. The backward loss output by the teacher and student networks which is trained by multiresolution knowledge distillation is applied to detect anomalous images [31], and similar to that, the STFPM [12] method directly uses the distillation residuals between the student and teacher convolutional neural networks to score for anomalies. Recently, the RD4AD [14] method detects anomalies by a reversed knowledge distillation with one-class embedding.

These great unsupervised methods for image anomaly detection take full use of normal images or normal pretrained features to fit the normal distribution and perform well, but they rarely apply limited labeled training anomalies to improve the detection accuracy. Also, most of the unsupervised methods rely on a threshold of the validation dataset to make an accurate binary detection.

B. Unbalanced Image Classification

Many research works on unbalanced image classification and related long-tailed learning have been studied recently [16]. Most of the unbalanced methods adopt the loss-adjustment-based approaches and perform well on unbalanced image classification.

The class-balanced loss [32] enforces a class-balanced reweighting term which is inversely proportional to the effective number of classes to address unbalanced problem. The FL [18] applies classifying scores to inversely reweight classes to indirectly address the unbalanced problem. A self-paced knowledge distillation framework [33] is introduced to learn a student network from multiple expert networks which are trained based on the subsets, to address unbalanced image classification. The influence-based classwise reweighting cross-entropy loss is proposed to address the overfitting of the majority classes in unbalanced image classification [19]. Distribution alignment between model prediction and balanced class prior is proposed to calibrate classification scores for unbalanced image classification [20]. Label-distribution-aware margin loss [34] enforces class-dependent margins based on label frequencies and encourages rare classes to have larger margins to handle class unbalance problem. The Logit adjustment loss (LAL) [21] adopts to adjust logit margin based on label frequencies, which encourages a large relative margin between logits of rare classes versus dominant classes, to address unbalanced image classification, but still cannot directly solve unbalanced anomalous image detection. The encouraging loss (EL) [35] is also studied recently to use the

well-classified images to continually optimize the deep model and performs well on unbalanced classification. An anomaly constrained loss is introduced to deep support vector data description [36] to drive a more discriminative classifying boundary to detect anomalies [15]. A mixed supervision training method is used to train a neural network to detect image anomalies out [5], but pixel-level labeled information is required. An asymmetric loss [37] developed from FL [18] is studied to give much more weight to the loss of anomalous images to train a neural network for image anomaly detection.

Most of the above methods perform well on unbalanced classification of natural images, but require an improvement for addressing unbalance detection of semi-supervised image anomaly detection, because labeled training anomalous images are limited and the number of them is much smaller than the one of normal images.

Thus, to address semi-supervised image anomaly detection, the LIAC method is proposed, in which an LIS is proposed to train an binary classifier and an ACM is proposed to learn discriminative feature representation for distinguishing anomalies.

III. Proposed LIAC Method

The proposed LIAC method for semi-supervised image anomaly detection is introduced in detail in this section.

For the problem of semi-supervised image anomaly detection, given an image $\boldsymbol{x} \in \mathbb{R}^{H\times W\times C}$ and its label $y \in \mathbb{R}^1$, where $y \in \{0, 1\}$, $y = 0$ denotes a normal image (negative sample) and $y = 1$ denotes an anomalous image (positive sample), and H, W, and C, respectively, denote the height, width, and channels. During the training stage, the number N_1 of anomalous images is limited, and much smaller than the number N_0 of normal images. Thus, it leads to an unbalanced binary detection problem of image anomaly detection, in which the label frequency p_1 of anomalous images is much smaller than the label frequency p_o of normal images. This task is aimed to learn a classifier $f_{\boldsymbol{\theta}}(\cdot)$ to detect anomalous images. Softmax normalization is adopted to transform the logit $f_{\boldsymbol{\theta}}(\cdot)[y]$ to represent the probability of $p_{\boldsymbol{\theta}}(\cdot)[y]$ as following:

$$p_{\boldsymbol{\theta}}(\cdot)[y] = \frac{\exp(f_{\boldsymbol{\theta}}(\cdot)[y])}{\sum_{y'} exp(\mathrm{f}_{\boldsymbol{\theta}}(\cdot)[\mathrm{y}'])} \tag{1}$$

where $\exp(\cdot)$ denotes the exponential function, and $y' \in \{0, 1\}$.

There are two problems required to be addressed, which, respectively, are unbalanced detection and anomaly detection. First, during testing, the high recall and accuracy of binary detection $\widehat{y}$ must be ensured though the labeled anomalies are limited during training

$$\widehat{y} = \begin{cases} 1, & p_{\boldsymbol{\theta}}(\cdot)[y] > 0.5 \\ 0, & p_{\boldsymbol{\theta}}(\cdot)[y] \le 0.5 \end{cases}. \tag{2}$$

The binary classifier $f_{\boldsymbol{\theta}}(\cdot)$ should not be biased toward the much more training normal images for detection. Second, the highly discriminative feature representation $g_{\boldsymbol{\varphi}}(\cdot)$ should be learned for both normal and anomalous images. Suppose there is a monotonically decreasing distance metric function $D(\cdot)$ and its monotonically increasing form invD$(\cdot)$, so that the highly discriminative features are optimized by $\ell_{\boldsymbol{\varphi}}$ objective function in the distance space

$$\ell_{\boldsymbol{\varphi}} = \sum_{i=1}^{N_0} D\left(g_{\boldsymbol{\varphi}}\left(\boldsymbol{x}_0^{(i)}\right)\right) + \sum_{j=1}^{N_1} \mathrm{invD}\left(g_{\boldsymbol{\varphi}}\left(\boldsymbol{x}_1^{(j)}\right)\right). \tag{3}$$

In another way, it means that the feature representations for anomalies are far away from the ones of the normal images, which achieved by the optimized $g_{\boldsymbol{\varphi}}(\cdot)$.

Therefore, the proposed LIAC method trains a binary classifier $f_{\boldsymbol{\theta}}(\cdot)$ with LIS to deal with the unbalanced detection problem and uses an ACM to learn high discriminative feature representation $g_{\boldsymbol{\varphi}}(\cdot)$ for distinguishing anomalies. The framework of the proposed LIAC for semi-supervised image anomaly detection is shown in Fig. 2.

A. Logit Inducing Loss

In semi-supervised image anomaly detection, the first problem must be addressed is the unbalanced problem; Otherwise, the learned binary classifier $f_{\boldsymbol{\theta}}(\cdot)$ is much biased toward normal images.

The loss-adjustment-based methods have shown well performance on unbalanced classification [16]. The long-standing baseline method WCE loss [17] based on label frequencies is introduced to address the class unbalance problem. Recently, some improved loss objective functions such as class-balanced loss [32], influence-based reweighting loss [19], and LAL [21] have been proposed to address unbalanced image classification. Inspired by the LAL [21] and EL [35], the LIL ℓ_{LIL} for optimizing binary classifier is proposed as following:

$$\begin{aligned} &\ell_{\mathrm{LIL}}(\boldsymbol{x}, y) \\ &\quad = \log\left(1 - y + \frac{p_{1-y}}{p_y}\exp(f_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] - f_{\boldsymbol{\theta}}(\boldsymbol{x})[y])\right) \\ &\quad = \log\left(1 - y + \exp\left(f_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] \right.\right. \\ &\qquad \left.\left. - f_{\boldsymbol{\theta}}(\boldsymbol{x})[y] + \log\left(\frac{p_{1-y}}{p_y}\right)\right)\right) \end{aligned} \tag{4}$$

where p_y denotes the label frequency of class y, $\log(\cdot)$ denotes the logarithmic function, and $\exp(\cdot)$ denotes the exponential function

$$\ell_{\mathrm{CEL}}(\boldsymbol{x}, y) = \log(1 + \exp(f_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] - f_{\boldsymbol{\theta}}(\boldsymbol{x})[y])). \tag{5}$$

Thus, compared with the vanilla cross-entropy loss ℓ_{CEL} for classification, the proposed ℓ_{LIL} requires that the logit $f_{\boldsymbol{\theta}}(\boldsymbol{x})[y]$ of the true label y should be larger than the wrong logit $f_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y]$ with a margin of $\log((p_{1-y}/p_y))$. When $\boldsymbol{x}$ belongs to the limited training anomalous images and $y = 1$, the label frequency p_1 is much smaller than the one p_0 of the dominant normal images of $y = 0$, and the margin term $\log((p_0/p_1))$ is a big constant term which drives that the optimized direction also focuses on the limited labeled anomalies during training. The unbalanced problem can be addressed to some extent, but the gradients of normal images dominantly drive the optimized direction during batchwise training since the labeled anomalies are much limited. On other hand, the proposed LIL also deals

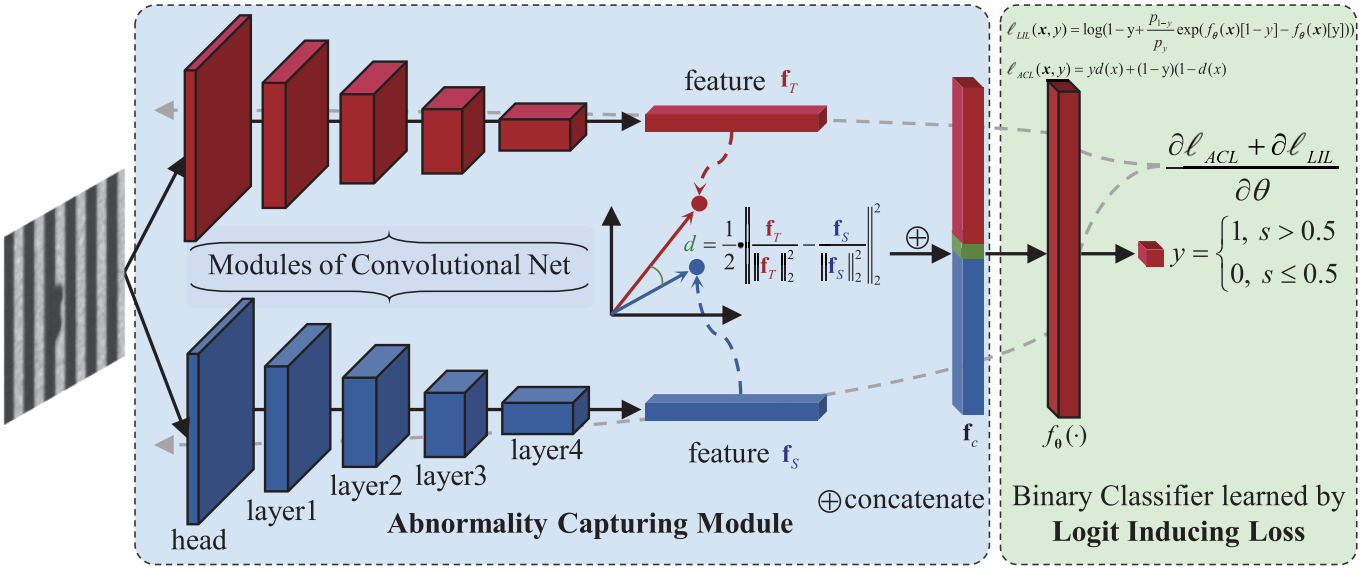


Fig. 2. Framework of the proposed LIAC method for semi-supervised image anomaly detection. The LIAC method consists of a binary classifier and an ACM, which, respectively, makes a binary anomaly detection and learns highly discriminative feature representation.

with the unbalanced problem by the backward-propagation gradients

$$\frac{\partial \ell_{\mathrm{LIL}}}{\partial \boldsymbol{\theta}} = \frac{\exp\left(f'_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] - f'_{\boldsymbol{\theta}}(\boldsymbol{x})[y] + \log\left(\frac{p_{1-y}}{p_y}\right)\right)}{1 - y + \Delta_y} \tag{6}$$

$$\Delta_y = \exp\left(f_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] - f_{\boldsymbol{\theta}}(\boldsymbol{x})[y] + \log\left(\frac{p_{1-y}}{p_y}\right)\right) \tag{7}$$

$$\frac{\partial \ell_{\mathrm{CEL}}}{\partial \boldsymbol{\theta}} = \frac{\exp\left(f'_{\boldsymbol{\theta}}(\boldsymbol{x})[1-y] - f'_{\boldsymbol{\theta}}(\boldsymbol{x})[y]\right)}{1 + \Delta_y} \tag{8}$$

where in (6), $f'_{\boldsymbol{\theta}}$ denotes the gradient $(\partial f_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta})$. Therefore, compared with the vanilla gradient of $(\partial \ell_{\mathrm{CEL}}/\partial \boldsymbol{\theta})$ which is described in (8), the gradient of $(\partial \ell_{\mathrm{LIL}}/\partial \boldsymbol{\theta})$ also continually focuses on limited labeled anomalies to optimize the model. This is because the gradient $(\partial \ell_{\mathrm{LIL}}/\partial \boldsymbol{\theta})|_{y=1}$ of labeled anomalies is approximately $(1 + \Delta_1/\Delta_1)$ times than the vanilla gradient of $(\partial \ell_{\mathrm{CEL}}/\partial \boldsymbol{\theta})|_{y=1}$. But the gradient $(\partial \ell_{\mathrm{LIL}}/\partial \boldsymbol{\theta})|_{y=0}$ of labeled normal images is approximately same. During batch-wise training, the term of Δ_1 of the well-classified anomalies approximately equals 0, so the ratio of $(1 + \Delta_1/\Delta_1)$ is a big term, which induces the gradients of anomalies to continually optimize the parameters of the binary classifier.

Therefore, in two aspects which are, respectively, classification margin adjustment and backward-propagation gradient adjustment, the proposed LIL can handle unbalanced detection by taking full use of limited labeled anomalies to learn a much discriminative binary classifier.

B. Abnormality Capturing Module

Although unbalanced detection can be addressed to some extent by the proposed LIS to train a binary classifier, the feature representation learned by the limited labeled anomalies is not enough because these anomalies cannot represent the whole anomalies. It is important to take full use of the dominant normal images, such as the unsupervised methods do [7], [11], [12], [13], to construct a powerful feature representation $g_{\boldsymbol{\varphi}}(\cdot)$ as described in (3).

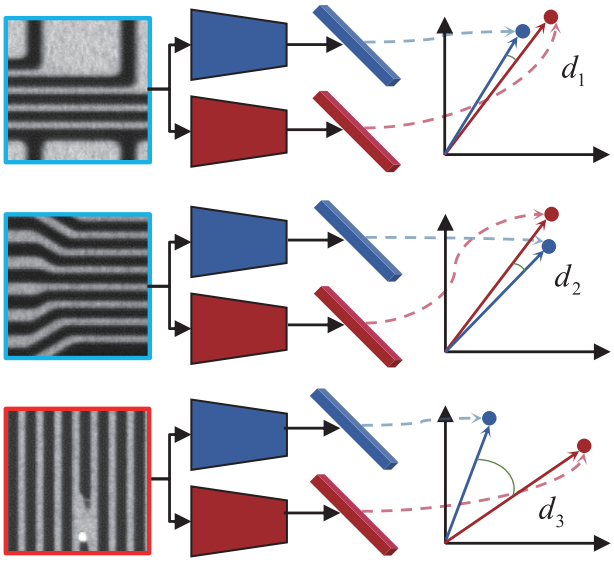


Fig. 3. Visualized description of the proposed ACM for feature learning. The ACM applies a couple of convolutional neural networks to extract a pair of features of an image, minimizing the distance between a pair of features of normal images, and maximizing the distance between pair of features of anomalous images during training. Images of the top two circled in light blue color are normal images, and the bottom image circled in light red color is an anomalous image.

To address this problem, the ACM is additionally proposed inspired by the work of knowledge distillation [12], [16], [31]. The ACM takes a couple of convolutional neural networks to learn the similarity of normal images and the dissimilarity of anomalous images as shown in Fig. 3. Therefore, both the dominant normal images and the limited anomalous images can be fully used for feature representation learning during training. The couple of convolutional neural networks share

the same designed structure and can be adopted as the popular backbone networks, such as VGG [38] or ResNet [39]. The layer of feature extracted by the convolutional neural network is defined as the final average pooling layer. The couple of convolutional neural networks are, respectively, defined as g_{φ_t} and g_{φ_s}, and the pair of features $\mathbf{f}_T, \mathbf{f}_S$ of an image $\boldsymbol{x}$ extracted by them is defined as following:

$$\mathbf{f}_T = g_{\varphi_t}(\boldsymbol{x}), \mathbf{f}_S = g_{\varphi_s}(\boldsymbol{x}). \tag{9}$$

For achieving similarity learning of normal images for two networks g_{φ_t} and g_{φ_s}, and dissimilarity learning of anomalous images, the distance metric function $D(\cdot)$ is defined as following:

$$D(\boldsymbol{x}) = \frac{1}{2}\| \frac{\mathbf{f}_T}{\|\mathbf{f}_T\|_2^2} - \frac{\mathbf{f}_S}{\|\mathbf{f}_S\|_2^2} \|_2^2. \tag{10}$$

Then, an abnormality capturing loss (ACL) ℓ_{ACL} is proposed in the ACM to learn the similarity and dissimilarity

$$\ell_{\text{ACL}} = yD(\boldsymbol{x}) + (1-y)(1-D(\boldsymbol{x})). \tag{11}$$

The ℓ_{ACL} loss helps learn the highly discriminative feature for distinguishing anomalies from normal images, and it is also beneficial for learning the binary classifier. Thus, the final feature $\mathbf{f}$ representation input to the binary classifier is a combination of all the features, and it is defined as following:

$$\mathbf{f} = [\mathbf{f}_T, \mathbf{f}_S, D(\boldsymbol{x})]. \tag{12}$$

Above all, to optimize the whole parameters $(\boldsymbol{\theta}, \boldsymbol{\varphi}_s, \boldsymbol{\varphi}_t)$ of the proposed LICA method, the total loss objective function is defined for training as following:

$$\ell = \ell_{\text{LIL}} + \ell_{\text{ACL}}. \tag{13}$$

Thus, the binary classifier and feature representation in the proposed LIAC method for semi-supervised image anomaly detection can be learned by optimizing the objective function ℓ.

IV. Experiments and Discussion

In this section, the datasets and the evaluation criteria are first introduced. Then the proposed LIAC method is compared with the state-of-the-art (SOTA) methods for image anomaly detection on all the datasets. More, the influence of ACL and of the number K of the labeled training anomalies is experimentally discussed.

A. Datasets and Evaluation Criteria

1) Printable Circuit Board Dataset: PCB is a self-collected PCB defect dataset from the production line of a cooperative manufacturing company. The anomalous image in the PCB dataset is caused by over, lack of tin solder, and other defect influence. The normal images in the PCB dataset can also change a lot in the texture structure, and the anomalies in the anomalous images widely range over size, shape, etc. These further improve the challenging to detect the anomalous images out. In Fig. 1, some samples of the PCB dataset are shown. The normal images are split into 6:2:2 for training, validation, and testing. The maximum of labeled anomalous images for training is set as 80, and the remaining are split into 5:5 for validation and testing.

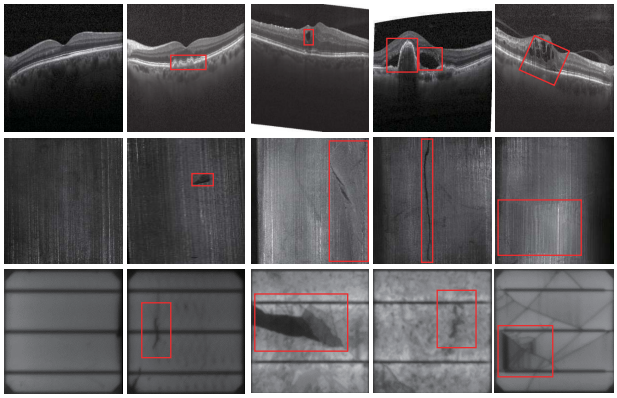


Fig. 4. Example images are, respectively, from the ROCT, MTD, and ELPV datasets. The top rows are from the ROCT dataset, the middle rows from the MTD dataset, and the bottom rows from the ELPV dataset. Anomalous regions are marked in red boxes in images and the first columns are normal images.

TABLE I

Numbers of Training, Validation, and Testing Sets of Each Dataset (normal:abnomalous)

Dataset	Training set	Validation set	Testing set
PCB	3000:(10-80)	1000:2740	1000:2740
MTD	571:(10-80)	190:154	191:158
ROCT	2000:(10-80)	242:726	242:726
ELPV	904:(10-80)	301:517	303:519

2) Retinal Optical Coherence Tomography Dataset: ROCT [23] is also an open-source dataset consisting of optical coherence tomography images of retina. The anomalous regions in images are caused by different diseases. Because the number of given normal images for training is too large, 2000 normal images are randomly chosen for training. The maximum of labeled anomalous images for training is also set as 80. The testing set has been split by ROCT, so the validation set of the same number is also randomly chosen from the remaining of the training set.

3) Magnetic Tile Defect Dataset: MTD [22] is also an open-source dataset consisting of 1344 images of the concerning surface of magnetic tiles. The pixel-labeled information is not used in this article. The anomalous image is difficult to detect since the number of normal images is small-scale compared with the PCB or ROCT dataset, and the anomalous regions are more fine-grained compared with the ones in the PCB or ROCT dataset. The splits of training, validation, and testing are same as the PCB dataset.

4) Electroluminescence Images of Photovoltaic Modules Dataset : ELPV [24] is also an open-source dataset consisting of defective and normal images of photovoltaic modules. There are some tiny and slight defects in the ELVP dataset, which are not obvious compared with normal images, and it is challenging to detect them out. The splits of training, validation, and testing are also the same as the PCB dataset.

The examples of images from the three datasets are shown in Fig. 4. The numbers of training, validation, and testing sets of four datasets are summarized in Table I, in which the term is the number of normal images: number of anomalous images. The expression of "10–80" denotes that the number

of labeled training anomalies is set from 10 to 80 for studying the influence of it on the supervised methods.

5) Evaluation Criteria: In this article, the accuracy Acc and F1-score $F1$ are adopted as the criteria of threshold-dependent to evaluate the performance. The accuracy criterion is biased to dominant images, but F1-score is fair for both. The area under the curve of precision–recall (AUCPR) and the area under the curve of the receiver-operating characteristic (AUCROC) are also adopted as the criteria of threshold-independent. For the supervised methods, $p_{\theta}(\cdot)$ [1] of the classified score is adopted to compute AUCPR and AUCROC, and 0.5 is adopted as the threshold to compute Acc and $F1$. For the compared unsupervised methods, the score output by them is adopted to compute AUCPR and AUCROC, and the threshold of maximum $F1$ of the precision–recall curve of the validation set is adopted as the threshold for the testing set to compute Acc and $F1$, which is fair for the unsupervised methods to compare with the criteria of threshold-dependent.

True positive (TP) is the number of detected anomalous images which are actually anomalies, false positive (FP) is the number of detected anomalous images which are actually normal images, true negative (TN) is the number of detected normal samples which are actually anomalies, and false negative (FN) is the number of detected normal samples which are actually normal images

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{17}$$

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{18}$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{19}$$

By scanning over the range of thresholds, sorted serial values of precision, recall, TP rate (TPR), and FP rate (FPR) are obtained and the AUCPR and AUCROC values can be calculated. precision_i, recall_i, TPR_i, and FPR_i, respectively, denote precision, recall, TPR, and FPR at the i sorted position, $d(\cdot)$ denotes the differential operation, and $T_{\max}$ is the maximum threshold; then AUCPR and AUCROC are denoted as following:

$$\text{AUCPR} = \sum_{i=1}^{T_{\max}} \text{precision}_i \cdot d(\text{recall}_i) \tag{20}$$

$$\text{AUCROC} = \sum_{i=1}^{T_{\max}} \text{TPR}_i \cdot d(\text{FPR}_i). \tag{21}$$

B. Implementation Details

The images are resized to 224×224 for input and normalized by the mean and standard deviation of the ImageNet dataset. For all the compared methods in this article, the ResNet34 convolutional neural network pretrained on the ImageNet dataset is adopted as the model which is light and appropriate for real-industry application. The SGD optimizer with an initial learning rate of $1e-3$, a batch size of 64, a total epoch of 100, and a learning rate halved at the 50th epoch are adopted for training. While for the compared methods, except for the ResNet34 backbone, the other settings are the same as their original ones. Python and PyTorch library are used to implement these methods. The hardware of Intel Xeon Gold 5118 CPU at 2.30 GHz with Nvidia Titan V GPU hardware is used for model training, validation, and testing. For training stage, random horizontal flipping is adopted as data augmentation.

In every experiment, all the results are averaged by random three independent runs. The model of maximum $F1$ on the validation set during training is stored for testing, and this is set for every method. For the supervised methods, the number of limited labeled anomalous images for training is set from 10 to 80, to study the influence of it.

C. Comparisons With the State-of-the-Art Methods

For comparing the proposed LIAC method with the SOTA methods, the SOTA unsupervised STFPM [12], ORTHOAD [30], MeanShift [40], CFlow [13], and RD4AD [14] are experimentally compared. Moreover, the strong baseline supervised methods WCE [17] and FL [18] and the SOTA supervised methods asymmetric loss (AsyL) [37], LAL [21], EL [35] are also experimentally compared. To fully compare the supervised methods with the unsupervised ones, the number K of the labeled training anomalies is set as 40 and 80 to, respectively, conduct the experiments. More, the results of the validation set and the testing set are calculated for comparing.

As shown in Table II, the applications of image anomaly detection on the PCB dataset are compared. Under the case that the number of labeled training anomalies is 40, the proposed LIAC method achieves higher performance than the SOTA methods in terms of the threshold-independent criteria of AUCROC and AUCPR. The threshold-dependent criteria of $F1$ and Acc of the unsupervised SOTA methods are evaluated by setting the threshold of the maximum $F1$ on the precision–recall curve, but when evaluating the proposed LIAC method and other supervised SOTA methods, the threshold is set as 0.5 for all the experiments. Even though the $F1$-score of the validation set is slightly lower than the best one of the RD4AD [14] method, the proposed LIAC method also achieves a comparing results compared with the SOTA methods in terms of the threshold-dependent criteria of $F1$ and Acc. Especially, the proposed LIAC method outperforms the supervised SOTA methods with a large margin. This is because the proposed LIAC method can address unbalanced detection well by the LIS, and also can learn highly discriminative feature representations by the ACM. Then, increasing the number of labeled training anomalies to 80, the proposed LIAC method outperforms the comparing SOTA methods in terms of all the criteria on both the validation and testing sets of PCB. It means that the LIAC method benefits from the increased training anomalies, which is superior than the unsupervised methods.

As shown in Table III, the proposed LIAC method is compared on the open-source MTD [22] dataset. In this dataset, the number of training normal images is small-scale

TABLE II

COMPARISON OF THE SOTA METHODS ON IMAGE ANOMALY DETECTION OF THE PCB DATASET. THE BEST ONE IS MARKED IN BOLD

	STFPM [12]	ORTHOAD [30]	MeanShift [40]	CFlow [13]	RD4AD [14]	WCE [17]	FL [18]	AsyL [37]	LAL [21]	EL [35]	LIAC
Results of validation set under 40 training anomalies.											
Acc	0.979	0.900	0.955	0.981	0.989	0.976	0.574	0.931	0.971	0.939	**0.996**
F1	0.986	0.936	0.969	0.987	**0.993**	0.956	0.557	0.886	0.947	0.898	0.992
AUCPR	0.999	0.946	0.994	0.999	1.000	0.999	0.995	0.999	0.999	0.999	**1.000**
AUCROC	0.997	0.905	0.985	0.997	0.999	0.997	0.985	0.998	0.997	0.998	**1.000**
Results of testing set under 40 training anomalies.											
Acc	0.983	0.900	0.947	0.959	0.989	0.974	0.580	0.926	0.972	0.938	**0.994**
F1	0.968	0.786	0.898	0.922	0.978	0.953	0.560	0.879	0.951	0.896	**0.988**
AUCPR	0.999	0.944	0.993	0.999	0.999	0.999	0.995	0.999	0.999	0.999	**1.000**
AUCROC	0.997	0.906	0.982	0.998	0.999	0.996	0.985	0.998	0.998	0.997	**1.000**
Results of validation set under 80 training anomalies.											
Acc	0.979	0.900	0.955	0.981	0.989	0.990	0.817	0.975	0.989	0.977	**0.998**
F1	0.986	0.936	0.969	0.987	0.978	0.981	0.745	0.955	0.979	0.960	**0.997**
AUCPR	0.999	0.946	0.994	0.999	0.999	**1.000**	0.999	**1.000**	**1.000**	**1.000**	**1.000**
AUCROC	0.997	0.905	0.985	0.997	**0.999**	**0.999**	0.996	**0.999**	**0.999**	**0.999**	**0.999**
Results of testing set under 80 training anomalies.											
Acc	0.983	0.900	0.947	0.959	0.989	0.987	0.805	0.973	0.985	0.976	**0.996**
F1	0.968	0.786	0.898	0.922	0.978	0.976	0.734	0.953	0.973	0.958	**0.992**
AUCPR	0.999	0.944	0.993	0.999	0.999	0.999	0.999	**1.000**	**1.000**	**1.000**	**1.000**
AUCROC	0.997	0.906	0.982	0.998	**0.999**	0.998	0.996	0.998	**0.999**	**0.999**	**0.999**

TABLE III

COMPARISON OF THE SOTA METHODS ON THE MTD DATASET. THE BEST ONE IS MARKED IN BOLD

	STFPM [12]	ORTHOAD [30]	MeanShift [40]	CFlow [13]	RD4AD [14]	WCE [17]	FL [18]	AsyL [37]	LAL [21]	EL [35]	LIAC
Results of validation set under 40 training anomalies.											
Acc	0.774	0.730	0.754	0.823	**0.879**	0.710	0.559	0.607	0.732	0.654	0.841
F1	0.759	0.707	0.761	0.805	0.865	0.787	0.715	0.737	0.791	0.761	**0.871**
AUCPR	0.799	0.718	0.763	0.871	**0.926**	0.823	0.604	0.827	0.802	0.845	0.923
AUCROC	0.821	0.761	0.820	0.863	**0.922**	0.837	0.630	0.834	0.818	0.848	0.919
Results of testing set under 40 training anomalies.											
Acc	0.736	0.668	0.756	0.763	0.813	0.703	0.549	0.603	0.714	0.634	**0.817**
F1	0.755	0.702	0.748	0.766	0.831	0.779	0.708	0.734	0.776	0.750	**0.852**
AUCPR	0.773	0.680	0.802	0.857	0.896	0.793	0.573	0.809	0.779	0.805	**0.910**
AUCROC	0.791	0.727	0.847	0.861	0.891	0.794	0.608	0.793	0.777	0.795	**0.901**
Results of validation set under 80 training anomalies.											
Acc	0.774	0.730	0.754	0.823	0.879	0.849	0.614	0.823	0.836	0.854	**0.949**
F1	0.759	0.707	0.761	0.805	0.865	0.878	0.741	0.861	0.867	0.882	**0.955**
AUCPR	0.799	0.718	0.763	0.871	0.926	0.941	0.794	0.937	0.923	0.947	**0.977**
AUCROC	0.821	0.761	0.820	0.863	0.922	0.942	0.810	0.939	0.927	0.944	**0.978**
Results of testing set under 80 training anomalies.											
Acc	0.736	0.668	0.756	0.763	0.813	0.820	0.601	0.814	0.825	0.809	**0.926**
F1	0.755	0.702	0.748	0.766	0.831	0.857	0.732	0.854	0.859	0.850	**0.936**
AUCPR	0.773	0.680	0.802	0.857	0.896	0.944	0.771	0.930	0.913	0.930	**0.967**
AUCROC	0.791	0.727	0.847	0.861	0.891	0.942	0.769	0.925	0.906	0.927	**0.963**

as shown in Table I, which increases the difficulty to make an accurate anomaly detection. Though the LIAC method achieves a slightly poor performance than the best one of RD4AD [14] method on the validation set under the case that the number of labeled training anomalies is 40, it outperforms all the comparing methods on the testing set. More, when increasing the number of labeled training anomalies to 80, the LIAC method achieves a better performance than the comparing SOTA methods with a large margin.

As shown in Table IV, the proposed LIAC method is also compared on the open-source ROCT [23] dataset. Under the two cases where the number of labeled training anomalies is, respectively, 40 and 80, though the $F1$-scores of the validation set are lower than the ones of the comparing RD4AD [14] method, it is also superior in all the criteria than all the comparing SOTA methods. It can be found that the proposed LIAC also improves as the number of the labeled training anomalies increases.

In addition, as shown in Table V, the proposed LIAC method is also compared on the open-source ELPV [24] dataset. In this dataset, the number of training normal images is also small-scale as shown in Table I, and there are many more slight defects in anomalous images which are similar to normal images, which further increases the challenge to make an accurate anomaly detection for all the methods. Under the two cases where the number of labeled training anomalies is, respectively, 40 and 80, though the $F1$- and Acc scores of the validation set are slightly lower than the ones of the comparing RD4AD [14] and CFlow [13] methods, the proposed LIAC method also outperforms all the comparing SOTA methods on the testing set in terms of threshold-independent and threshold-dependent criteria. This further shows the superiority of the LIAC method in image anomaly detection.

More, the visualization of the confusion matrix [41] of image anomaly detection is added to compare. As shown in Figs. 5 and 6, the visualizations on the PCB and MTD datasets are compared under the two cases where the number of labeled training anomalies is, respectively, 40 and 80. It can be viewed that the proposed LIAC method can accurately detect normal images compared with the unsupervised SOTA methods and

TABLE IV

COMPARISON OF THE SOTA METHODS ON THE ROCT DATASET. THE BEST ONE IS MARKED IN BOLD

	STFPM [12]	ORTHOAD [30]	MeanShift [40]	CFlow [13]	RD4AD [14]	WCE [17]	FL [18]	AsyL [37]	LAL [21]	EL [35]	LIAC
Results of validation set under 40 training anomalies.											
Acc	0.835	0.753	0.763	0.799	0.857	0.888	0.304	0.678	0.842	0.694	**0.905**
F1	0.898	0.859	0.863	0.876	**0.909**	0.812	0.418	0.609	0.743	0.620	0.834
AUCPR	0.934	0.783	0.914	0.918	0.957	0.989	0.962	0.979	0.978	0.979	**0.989**
AUCROC	0.845	0.568	0.772	0.795	0.891	**0.964**	0.888	0.933	0.931	0.933	0.962
Results of testing set under 40 training anomalies.											
Acc	0.915	0.753	0.769	0.879	0.961	0.977	0.360	0.833	0.941	0.860	**0.983**
F1	0.800	0.042	0.167	0.789	0.919	0.956	0.439	0.753	0.892	0.782	**0.968**
AUCPR	0.993	0.780	0.944	0.987	0.996	**0.999**	0.990	0.996	0.997	0.996	**0.999**
AUCROC	0.981	0.582	0.850	0.961	0.991	**0.997**	0.967	0.987	0.990	0.986	**0.997**
Results of validation set under 80 training anomalies.											
Acc	0.835	0.753	0.763	0.799	0.857	0.906	0.545	0.801	0.874	0.803	**0.934**
F1	0.898	0.859	0.863	0.876	**0.909**	0.837	0.524	0.715	0.788	0.717	0.878
AUCPR	0.934	0.783	0.914	0.918	0.957	0.991	0.977	0.987	0.984	0.985	**0.992**
AUCROC	0.845	0.568	0.772	0.795	0.891	0.970	0.928	0.957	0.948	0.951	**0.973**
Results of testing set under 80 training anomalies.											
Acc	0.915	0.753	0.769	0.879	0.961	0.982	0.680	0.938	0.969	0.950	**0.991**
F1	0.800	0.042	0.167	0.789	0.919	0.965	0.611	0.889	0.941	0.909	**0.983**
AUCPR	0.993	0.780	0.944	0.987	0.996	0.999	0.996	0.998	0.998	0.999	**1.000**
AUCROC	0.981	0.582	0.850	0.961	0.991	0.997	0.987	0.995	0.993	0.995	**0.999**

TABLE V

COMPARISON OF THE SOTA METHODS ON THE ELPV DATASET. THE BEST ONE IS MARKED IN BOLD

	STFPM [12]	ORTHOAD [30]	MeanShift [40]	CFlow [13]	RD4AD [14]	WCE [17]	FL [18]	AsyL [37]	LAL [21]	EL [35]	LIAC
Results of validation set under 40 training anomalies.											
Acc	0.669	0.651	0.675	0.677	**0.691**	0.655	0.398	0.492	0.680	0.515	0.670
F1	0.790	0.783	0.793	0.790	**0.800**	0.640	0.550	0.591	0.662	0.601	0.663
AUCPR	0.839	0.723	0.850	**0.881**	0.861	0.858	0.826	0.856	0.860	0.856	0.870
AUCROC	0.725	0.632	0.748	**0.780**	0.757	0.774	0.710	0.747	0.760	0.757	0.776
Results of testing set under 40 training anomalies.											
Acc	0.641	0.642	0.660	0.646	0.667	0.663	0.393	0.483	0.665	0.507	**0.678**
F1	0.168	0.095	0.223	0.117	0.276	0.634	0.548	0.587	0.642	0.596	**0.666**
AUCPR	0.844	0.731	0.834	0.877	0.855	0.850	0.818	0.856	0.852	0.852	**0.879**
AUCROC	0.704	0.644	0.721	0.769	0.732	0.771	0.711	0.749	0.750	0.753	**0.787**
Results of validation set under 80 training anomalies.											
Acc	0.669	0.651	0.675	0.677	**0.691**	0.675	0.469	0.585	0.667	0.564	0.674
F1	0.790	0.783	0.793	0.790	**0.800**	0.657	0.580	0.633	0.658	0.623	0.678
AUCPR	0.839	0.723	0.850	0.881	0.861	0.868	0.835	0.861	0.865	0.866	**0.879**
AUCROC	0.725	0.632	0.748	0.780	0.757	0.778	0.727	0.756	0.769	0.769	**0.785**
Results of testing set under 80 training anomalies.											
Acc	0.641	0.642	0.660	0.646	0.667	0.685	0.460	0.586	0.664	0.573	**0.687**
F1	0.168	0.095	0.223	0.117	0.276	0.662	0.577	0.632	0.654	0.629	**0.690**
AUCPR	0.844	0.731	0.834	0.877	0.855	0.869	0.842	0.864	0.866	0.874	**0.898**
AUCROC	0.704	0.644	0.721	0.769	0.732	0.772	0.733	0.764	0.765	0.777	**0.813**

can accurately detect anomalous images compared with the supervised SOTA methods. Such as under the number of labeled training anomalies is 40, the LIAC method always detects the true anomalies more accurately than the supervised methods whether for the PCB or MTD dataset, and it also ensures lower false detection when detecting normal images than the unsupervised methods. Especially, when increasing the number of labeled training anomalies to 80, the proposed LIAC method outperforms all the comparing SOTA methods from visualizations of the confusion matrix.

For comparing the processing time for an image to make anomaly detection, the testing time for each compared method is shown in Table VI. The supervised WCE [17], FL [18], AsyL [37], LAL [21], and EL [35] take less processing time to detect because they just rely on the single network for forward propagation. The proposed LIAC method also achieves a slightly faster processing time. This is because the LIAC method detects anomalies by a couple of networks for forward propagation.

To summarize, whether for the threshold-independent criteria of AUCPR and AUCROC or the threshold-dependent criteria of $F1$ and Acc, the proposed LIAC method achieves the comparing results compared with the unsupervised and supervised SOTA methods and outperforms them as increasing the number of the labeled training anomalies. More, it can be found that the LIAC achieves comparing $F1$-score on the testing set of four datasets compared with the SOTA methods, and this is important because higher $F1$ score means that LIAC accurately detects the anomalies while keeping a lower false detection rate. Totally, from the comparison on the PCB, MTD [22], ROCT [23], and ELPV [24] four datasets, the LIAC can both achieve a better performance and show a well-generalized ability on different datasets.

D. Influence of the Abnormality Capturing Loss

The proposed ACL is important to the proposed LIAC method, because it helps the ACM to learn highly discriminative feature representation by a couple of networks.

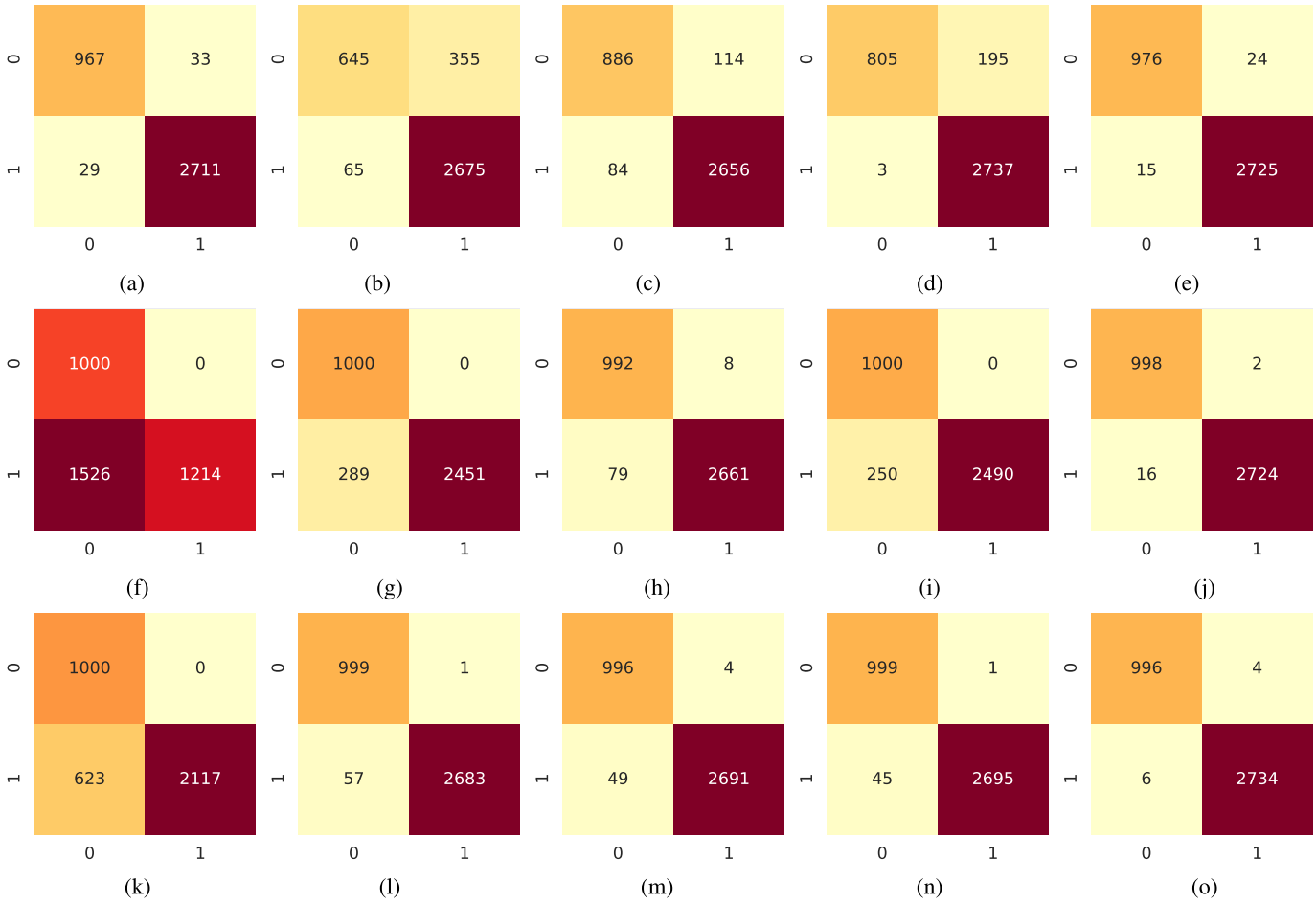


Fig. 5. Visualizations of the confusion matrix [41] on the testing set of PCB are output by each method. The number of labeled training anomalies is 80 in this case. The y-axis is the true label, and the x-axis is the predicting one. The label y of 1 denotes the anomalous image, and the label y of 0 denotes the normal image. The number after the name of supervised methods denotes the number of labeled anomalies for training. (a) STFPM [12]. (b) ORTHOAD [30]. (c) MeanShift [40]. (d) CFlow [13]. (e) RD4AD [14]. (f) FL_40 [18]. (g) AsyL_40 [37]. (h) LAL_40 [21]. (i) EL_40 [35]. (j) LIAC_40. (k) FL_80 [18]. (l) AsyL_80 [37]. (m) LAL_80 [21]. (n) EL_80 [35]. (o) LIAC_80.

TABLE VI
COMPARISON OF THE SOTA METHODS ON TESTING TIME FOR AN IMAGE (/MS)

	STFPM [12]	ORTHOAD [30]	MeanShift [40]	CFlow [13]	RD4AD [14]	WCE [17]	FL [18]	AsyL [37]	LAL [21]	EL [35]	LIAC
Time	1.80	4.64	4.04	9.72	23.1	1.27	1.27	1.27	1.27	1.27	1.75

The ACL is applied to optimize the similarity between normal images and the dissimilarity between anomalous images. Therefore, the dominant normal images can be adequately used like the unsupervised methods, to indirectly capture the anomalies that are far away from the distribution of normal images.

The experiments for studying the influence of ACL are conducted under each number K of training anomalies from 10 to 80 on four datasets. The results of the validation and testing sets of each dataset with or without ACL are shown in Fig. 7. It can be viewed that the ACL can improve the performance of the LIAC method a lot for image anomaly detection on the PCB and ROCT datasets, especially under the labeled training anomalies that are limited from 10 to 40. This is because the number of normal images in the PCB and ROCT datasets is relatively large, and the ACL can help the LIAC learn highly discriminative feature representation by dominant normal images. Though the number of normal images in the MTD and ELPV datasets is relatively small, there is also some improved performance achieved by the ACL. Anomalies in the two datasets diversely range in shape, size, etc., which further increases the difficulty to make a large improvement. In the ELPV dataset, there are many slight anomalies that are similar to normal images, and this also makes the detection challenging. Through these experiments, the ACL can improve the performance in most cases.

E. Influence of the Number K of Labeled Training Anomalies

In this section, the influence of the number K of labeled training anomalies on each method is experimentally studied. The number K is set from 10 to 80, and the stride is 10. For the unsupervised methods which are STFPM [12], ORTHOAD [30], MeanShift [40], CFlow [13],

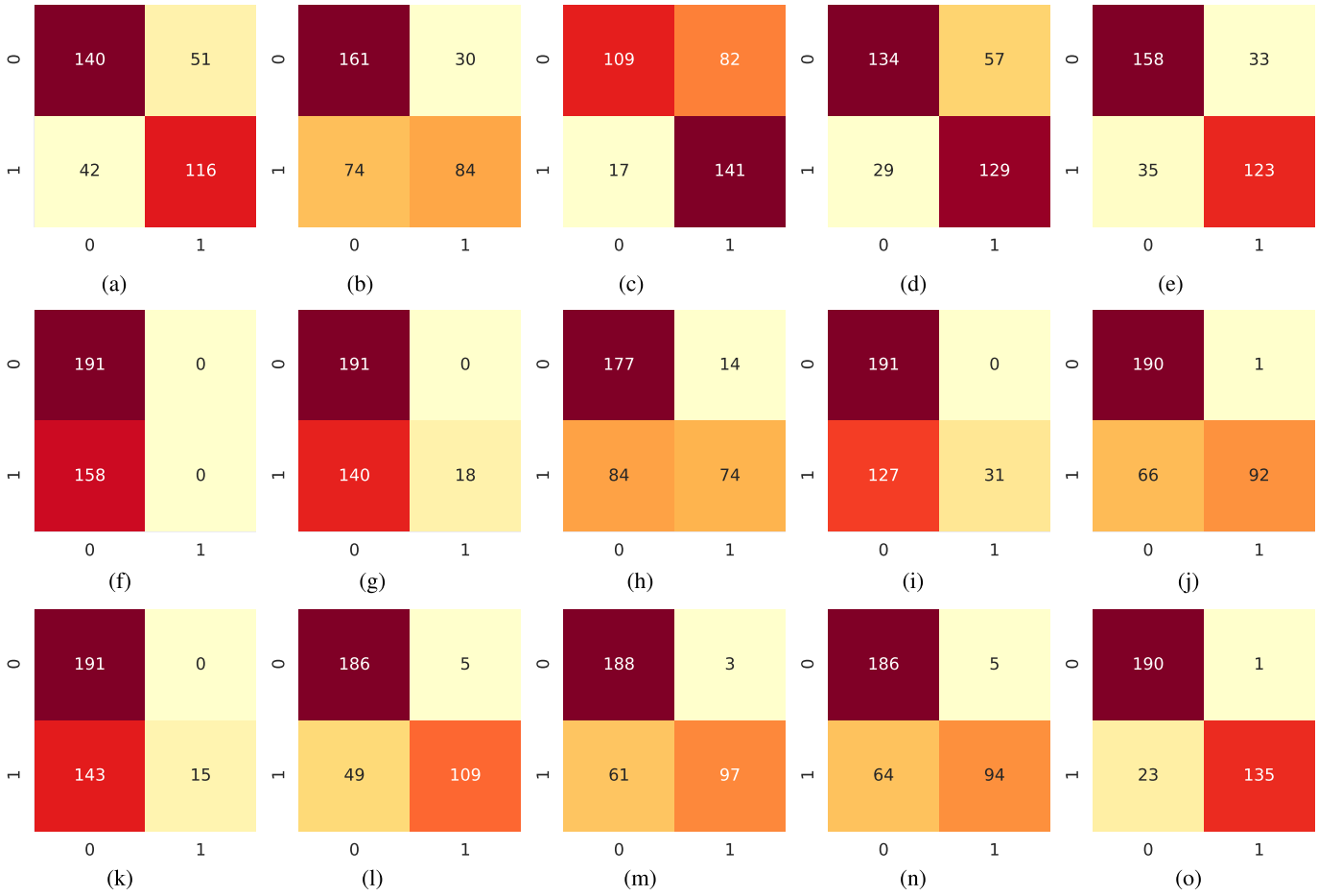


Fig. 6. Visualizations of the confusion matrix [41] on the testing set of MTD are output by each method. The number of labeled training anomalies is 80 in this case. The y-axis is the true label, and the x-axis is the predicting one. The label y of 1 denotes the anomalous image, and the label y of 0 denotes the normal image. The number after the name of supervised methods denotes the number of labeled anomalies for training. (a) STFPM [12]. (b) ORTHOAD [30]. (c) MeanShift [40]. (d) CFlow [13]. (e) RD4AD [14]. (f) FL_40 [18]. (g) AsyL_40 [37]. (h) LAL_40 [21]. (i) EL_40 [35]. (j) LIAC_40. (k) FL_80 [18]. (l) AsyL_80 [37]. (m) LAL_80 [21]. (n) EL_80 [35]. (o) LIAC_80.

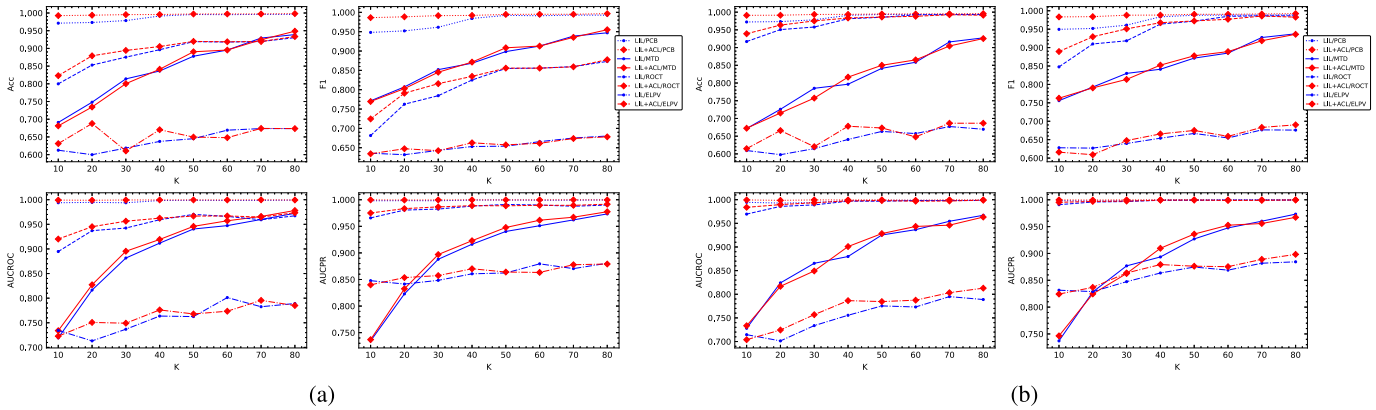


Fig. 7. Studying of the influence of the proposed ACL on each dataset. (a) Validation set. (b) Testing set.

and RD4AD [14], they only adopt normal images for training, and the number K of training anomalies does not affect their performance. For the remaining supervised methods, the experiments are conducted under each number K of training anomalies from 10 to 80 on four datasets.

As shown in Fig. 8, the results of testing set output by each method under each number K of training anomalies are reported. First for the PCB and ROCT datasets, it can be viewed that the LIAC method outperforms the comparing methods even though the labeled training anomalies are limited, such as 10 or 20. This is because the number of training normal images of the PCB or ROCT dataset is large, which is, respectively, 3000 and 2000, so that the ACL ℓ_{ACL} of the proposed LIAC method can take full use of these normal

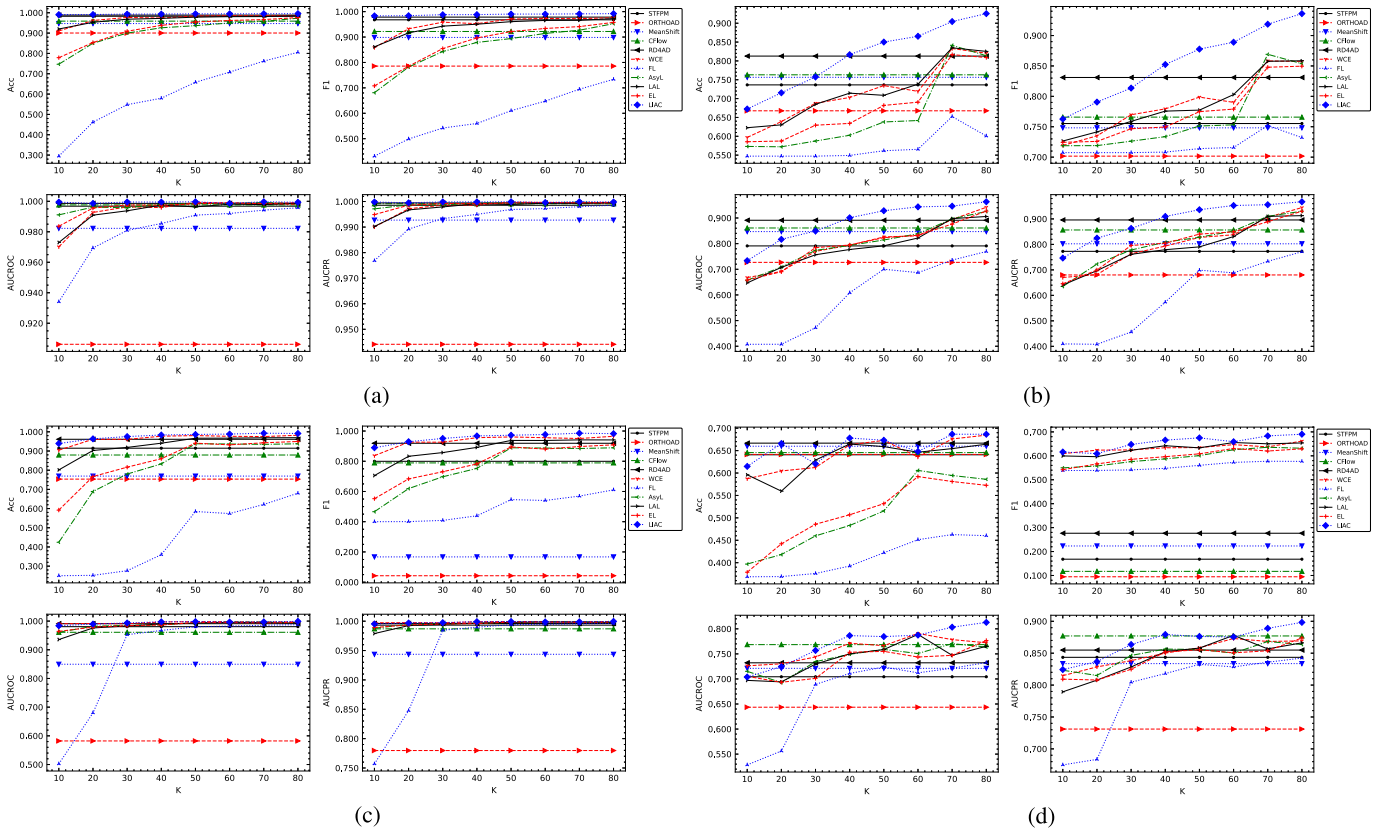


Fig. 8. Studying of the influence of the labeled training anomalies K for each methods on each dataset. (a) PCB. (b) MTD [22]. (c) ROCT [23]. (d) ELPV [24].

images to indirectly learn the dissimilarity of anomalies. Then for the MTD and ELPV datasets, it can be viewed that the LIAC method outperforms the comparing methods when the labeled training anomalies increased to 40. This is because the number of training normal images of them is relatively small compared with the PCB dataset, and also the anomalies in the two datasets are much more diverse. The above two problems increase the difficulty to make an accurate detection for the proposed LIAC method under a much small number of labeled training anomalies. When increasing the number of labeled training anomalies to 40 and more, the LIAC method can also outperform the comparing SOTA methods on the MTD and ELPV datasets.

All in all, the limited labeled training anomalies are important not only for the proposed LIAC method but also for all the supervised methods, which can help improve the detection performance in a margin.

V. Conclusion and Future Work

A novel LIAC method is proposed for image anomaly detection in semi-supervised manners, consisting of an LIS and an ACM, which, respectively, addresses the unbalanced problem and deals with the learning of highly discriminative feature representation. Extensive experiments on four datasets have shown the superiority of the LIAC method compared with the SOTA methods. The LIAC method achieves, respectively, 98.8%, 85.2%, 96.8%, and 66.6% of $F1$-score on the PCB, MTD, ROCT, and ELPV datasets under only given 40 labeled training anomalies. The performance is competing compared with the SOTA methods. It also achieves comparing testing time for detection with 1.75 ms per image. But the LIAC still performs not well on some scenarios with limited labeled training anomalies, especially when the normal images and anomalous images are much more diverse, fine-grained, and they are unknown to the model. In future work, decoupled learning can be borrowed to address more discriminative feature learning to solve fine-grained and unknown anomaly detection.

References

[1] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 626–644, Jan. 2020.

[2] Z. Zhao *et al.*, "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–28, 2021.

[3] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, May 2020.

[4] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

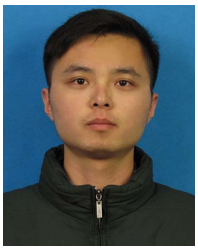
[5] J. Božič, D. Tabernik, and D. Skočaj, "Mixed supervision for surface-defect detection: From weakly to fully supervised learning," *Comput. Ind.*, vol. 129, Jan. 2021, Art. no. 103459.

[6] L. Ruff *et al.*, "A unifying review of deep and shallow anomaly detection," *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, May 2021.

[7] Q. Wan, L. Gao, X. Li, and L. Wen, "Industrial image anomaly localization based on Gaussian clustering of pretrained feature," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 6182–6192, Jul. 2022.

[8] V. Zavrtanik, M. Kristan, and D. Skocaj, "DRAEM—A discriminatively trained reconstruction embedding for surface anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8310–8319.

[9] Y. Shi, J. Yang, and Z. Qi, "Unsupervised anomaly segmentation via deep feature reconstruction," *Neurocomputing*, vol. 424, pp. 9–22, Feb. 2021.
[10] Y. Cao, Q. Wan, W. Shen, and L. Gao, "Informative knowledge distillation for image anomaly segmentation," *Knowl.-Based Syst.*, vol. 248, Jul. 2022, Art. no. 108846.
[11] Q. Wan, L. Gao, X. Li, and L. Wen, "Unsupervised image anomaly detection and segmentation based on pre-trained feature mapping," *IEEE Trans. Ind. Informat.*, early access, Jun. 13, 2022, doi: 10.1109/TII.2022.3182385.
[12] G. Wang, S. Han, E. Ding, and D. Huang, "Student-teacher feature pyramid matching for anomaly detection," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2021, pp. 1–14.
[13] D. Gudovskiy, S. Ishizaka, and K. Kozuka, "CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1819–1828.
[14] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9737–9746.
[15] L. Ruff *et al.*, "Deep semi-supervised anomaly detection," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–23.
[16] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," 2021, *arXiv:2110.04596*.
[17] Y. Xie and C. F. Manski, "The logit model and response-based samples," *Sociol. Methods Res.*, vol. 17, no. 3, pp. 283–302, 1989.
[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. ICCV*, 2017, pp. 2980–2988.
[19] S. Park, J. Lim, Y. Jeon, and J. Y. Choi, "Influence-balanced loss for imbalanced visual classification," in *Proc. ICCV*, 2021, pp. 735–744.
[20] S. Zhang, Z. Li, S. Yan, X. He, and J. Sun, "Distribution alignment: A unified framework for long-tail visual recognition," in *Proc. CVPR*, Jun. 2021, pp. 2361–2370.
[21] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–27.
[22] Y. Huang, C. Qiu, Y. Guo, X. Wang, and K. Yuan, "Surface defect saliency of magnetic tile," in *Proc. IEEE 14th Int. Conf. Autom. Sci. Eng. (CASE)*, 2018, pp. 612–617.
[23] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
[24] S. Deitsch *et al.*, "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Sol. Energy*, vol. 185, pp. 455–468, Jun. 2019.
[25] B. Staar, M. Lütjen, and M. Freitag, "Anomaly detection with convolutional neural networks for industrial surface inspection," *Proc. CIRP*, vol. 79, pp. 484–489, Jan. 2019.
[26] J. Yi and S. Yoon, "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation," in *Computer Vision—ACCV*. Cham, Switzerland: Springer, 2021, pp. 375–390.
[27] S. Venkataramanan, K.-C. Peng, R. V. Singh, and A. Mahalanobis, "Attention guided anomaly localization in images," in *Computer Vision*. Cham, Switzerland: Springer, 2020, pp. 485–503.
[28] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proc. CVPR*, Jun. 2020, pp. 4182–4191.
[29] O. Rippel, P. Mertens, E. König, and D. Merhof, "Gaussian anomaly detection by modeling the distribution of normal data in pretrained deep features," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
[30] J. Kim, D. Kim, S. Yi, and T. Lee, "Semi-orthogonal embedding for efficient unsupervised anomaly segmentation," 2021, *arXiv:2105.14737*.
[31] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proc. CVPR*, Jun. 2021, pp. 14897–14907.
[32] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proc. CVPR*, Jun. 2019, pp. 9268–9277.
[33] L. Xiang, G. Ding, and J. Han, "Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2020, pp. 247–263.
[34] K. Cao, C. Wei, A. Gaidon, N. Aréchiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," in *Advances in Neural Information Processing Systems*, H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, Eds. Vancouver, BC, Canada: Curran Associates, 2019, pp. 1565–1576.
[35] G. Zhao, W. Yang, X. Ren, L. Li, and X. Sun, "Well-classified examples are underestimated in classification with deep neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 1–10.
[36] L. Ruff *et al.*, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
[37] Z. Lin, H. Ye, B. Zhan, and X. Huang, "An efficient network for surface defect detection," *Appl. Sci.*, vol. 10, no. 17, p. 6085, 2020.
[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–14.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Jun. 2016, pp. 770–778.
[40] T. Reiss and Y. Hoshen, "Mean-shifted contrastive loss for anomaly detection," 2021, *arXiv:2106.03844*.
[41] A. Theissler, M. Thomas, M. Burch, and F. Gerschner, "Confusion-Vis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices," *Knowl.-Based Syst.*, vol. 247, Jul. 2022, Art. no. 108651.



Qian Wan was born in Jiangxi, China, in 1996. He received the B.S. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2018, where he is currently pursuing the Ph.D. degree in mechanical and electronic engineering.

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

Liang Gao (Senior Member, IEEE) 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 the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2002.

He is currently a Professor with the Department of Industrial and Manufacturing System Engineering, the Deputy Director of the State Key Laboratory of Digital Manufacturing Equipment and Technology, and the Vice Dean of the 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. He has authored more than 200 refereed articles indexed by SCIE and has authored seven monographs. His research interests include operations research and optimization, big data, and machine learning.

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

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

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