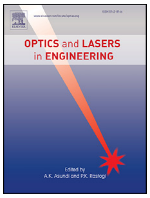
Optics and Lasers in Engineering 136 (2021) 106324

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng



Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection

Jie Liu[a,b], Kechen Song[a,b,*], Mingzheng Feng[a,b], Yunhui Yan[a,b,*], Zhibiao Tu[c], Liu Zhu[c]

[a] School of Mechanical Engineering & Automation, Northeastern University, Shenyang, Liaoning, China
[b] Key Laboratory of Vibration and Control of Aero-Propulsion Systems Ministry of Education of China, Northeastern University, Shenyang, Liaoning, China
[c] School of Pharmaceutical and Materials Engineering, Taizhou University, Taizhou, Zhejiang, China

ARTICLE INFO

Keywords:
Anomaly detection
Auto-encoder
Surface defect
Semi-supervised learning

ABSTRACT

Anomaly detection in the automated optical quality inspection is of great important for guaranteeing the surface quality of industrial products. Most related methods are based on supervised learning techniques, which require a large number of normal and anomalous samples to obtain a robust classifier. However, the diversity of potential defects and low availability of defective samples during manufacturing bring more challenges to anomaly detection. Based on the encoder-decoder-encoder paradigm, a semi-supervised anomaly detection method Dual Prototype Auto-Encoder (DPAE) is proposed in this paper. At the training stage, the dual prototype loss and reconstruction loss are introduced to encourage the latent vectors generated by the encoders to keep closer to their own prototype. Therefore, two latent vectors of the normal image tend to be closer, and large distance between the latent vectors indicates an anomaly. And we also construct the Aluminum Profile Surface Defect (APSD) dataset for the anomaly detection task. Finally, extensive experiments on four datasets show that DPAE is effective and outperforms state-of-the-art methods.

1. Introduction

Automated optical quality inspection is a crucial method for guaranteeing the surface quality of industrial products. Compared with manual detection, vision-based methods exhibit superiority in satisfying the intelligent and high-efficiency demands of the modern industrialized manufacturing line. Therefore, many vision-based models have been proposed to inspect the surface quality of industrial products, such as aluminum profile surfaces [4,35], steel surfaces [1,5,16,18,36,42], wooden surfaces [19] and textured fabric [20,21].

A more fundamental and practical task in the optical industrial quality inspection is anomaly detection [29], which involves discovering of divergence of defective (anomalous) samples from defect-free (normal) samples. Anomaly detection models can be broadly classified into three categories based on the availability of labels [38], i.e., supervised deep anomaly detection, semi-supervised anomaly detection, and unsupervised anomaly detection. With abundant labeled defective and defect-free samples, the anomaly detection task can be addressed as a supervised binary classification task [22,23]. However, industrial processes are always optimized to minimize the anomaly accident of industrial production, resulting in a minimal number of defective samples and a vast number of images without defects. Therefore, semi-supervised anomaly detection methods [10,12,16,39], the aim of which is training merely on defect-free samples and then accurately distinguishing both defective and defect-free samples, is more realistic in the industrial scenario. In addition, when there are a lot of unlabeled training samples, the model trained on these samples can be regarded as unsupervised anomaly detection method [3,8,11,13,34], which intends to learn a hyperplane that can separate defective and defect-free samples accurately.

In fact, anomaly detection of industrial production surfaces faces many challenges. Firstly, in the complicated industrial scenario, some uncertain factors and different production processes lead to the difference between defect-free samples. Therefore, a limited amount of training data cannot represent the distribution of all defect-free images, resulting in the ambiguity between unseen defect-free samples and anomalies. Secondly, the robustness of the model learned on defect-free samples is restricted by the diversity of defects. It is usually unpredictable what types of defects might appear in the manufacturing, and different defects often exhibit great inter-class differences. Finally, high similarity between defect-free and defective samples brings more difficulties to the recognition between the normal and anomalous. As shown in Fig. 1(c), the boundary between normal and anomalous is ambiguous. Therefore, a robust anomaly detection model is required to make full use of the information of defect-free samples to distinguish defective samples.

To address the above challenges, we propose a novel Dual Prototype Auto-Encoder, namely DPAE, for anomaly detection, which achieves

[*] Corresponding authors.
E-mail addresses: songkc@me.neu.edu.cn (K. Song), yanyh@mail.neu.edu.cn (Y. Yan).

https://doi.org/10.1016/j.optlaseng.2020.106324
Received 20 March 2020; Received in revised form 11 July 2020; Accepted 30 July 2020

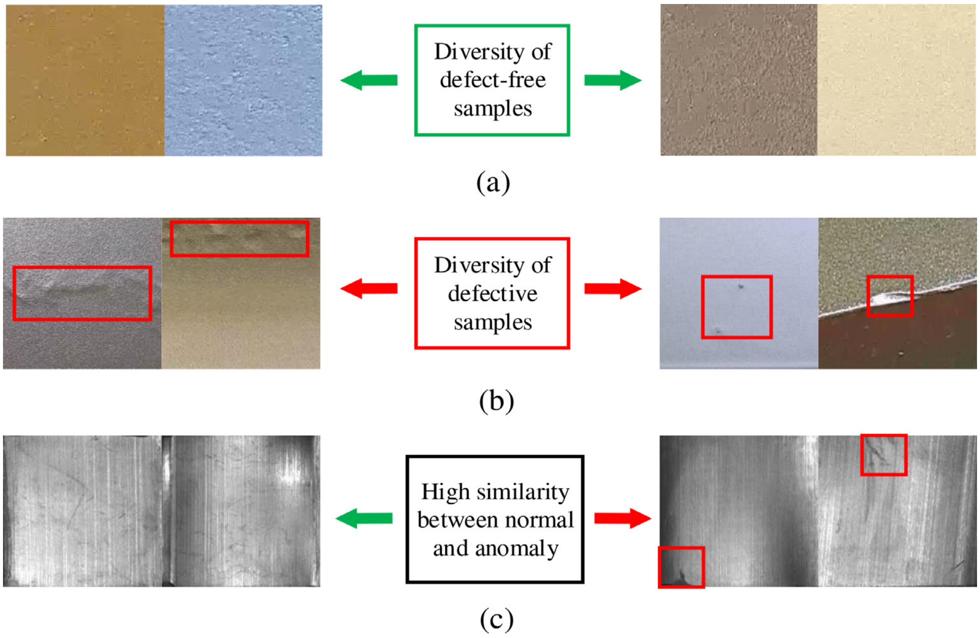

Diversity of defect-free samples
(a)
Diversity of defective samples
(b)
High similarity between normal and anomaly
(c)

Fig. 1. Challenges of anomaly detection of industrial production surfaces. (a) Diversity of defect-free samples. (b) Diversity of defective samples. (c) High similarity between defect-free and defective samples.

promising results on four datasets. As DPAE is a semi-supervised method, which requires merely defect-free samples at the training stage and aims to recognize both defect-free and defective samples at the test stage, it can mitigate or avoid the influence of the low availability of defective samples. To minimize the impact of diversity of defect-free and defective samples, DPAE follows an encoder-decoder-encoder paradigm. And the dual prototype loss is therefore proposed to encourage the output of each encoder, i.e., the low-dimension latent vectors, to keep closer to their respective prototypes, which represent the center of the latent vectors of defect-free samples. The prototypes are initialized as random vectors and updated with the training process. Besides, the reconstructed loss between the training samples and their reconstructed version (the output of the decoder) is adopted to distinguish the defective samples that share high similarity with defect-free samples. Finally, the mean square error between the low-dimension latent vectors can be utilized as an indicator of anomalies.

Specifically, the main contributions of this paper are summarized as follows:

- We propose a semi-supervised anomaly detection method for industrial product surface images, and it requires no anomalous training images and can be trained in an end-to-end manner.
- The dual prototype loss is introduced to encourage the latent vectors generated by the encoders to keep closer to their own prototype. Therefore, the mean square error between the latent vectors can be used as an indicator of anomalies.
- We construct the Aluminum Profile Surface Defect (APSD) dataset with defective and defect-free images to evaluate the performance of anomaly detection methods comprehensively in the industrial scenario.
- Extensive experiments are conducted on both our constructed dataset and the other three defect datasets, and the results demonstrate the effectiveness of our proposed method in the industrial scenario where defective samples are usually unavailable.

2. Related work

2.1. Anomaly detection

Anomaly detection techniques are widely used in many fields, such as video surveillance [28], medical diagnostics [11,25,26], surface defect detection [27,30,31], and credit card fraud detection [32]. Besides, many works have been published to summarize these approaches in the literatures [8,17,26,38].

However, anomaly detection on image data is still challenging. Domain-based anomaly detection models are classical approaches, which attempt to learn a discriminative boundary around normal data, such as one-class SVM [12], kernel density estimation (KDE) [6]. And some methods based on unsupervised clustering, such as fuzzy c-means clustering [24] and Gaussian Mixture Models (GMM) [8,9], have also been used to model the distribution of the normal samples to recognize anomaly. However, these methods show poor performance when applied to high-dimension data like images.

Recently, reconstruction-based methods have exhibited remarkable performance on the image anomaly detection task [2,10,11,13,41]. GANomaly [10] proposes to complete anomaly detection task by adversarial training. The deep convolutional generative adversarial network is utilized in AnoGAN [11] to find anomaly on both image and pixel level. Memory-augmented autoencoder is proposed in [2] to mitigate the problem that sometimes autoencoder "generalizes" so well that it can reconstruct anomalies well. [13] uses Variational auto-encoder (VAE) to localize the anomalous (tumor) region of the brain images. A robust reconstruction-based model is required to induce larger reconstruction error on the anomalies, in contrast to lower reconstruction error on normal samples.

2.2. Auto-Encoder

Autoencoder (AE) is a powerful tool to model the high-dimensional data in the unsupervised setting [2]. An autoencoder is comprised of an encoder and a decoder, the encoder compresses the input data into a latent representation, and the decoder maps the latent representation back to the original input space. It is trained to reconstruct the input data as much as possible. In the anomaly detection task, AE is usually trained by minimizing the reconstruction error of defect-free samples, and then the reconstruction error is adopted as an indicator of anomalies. Generally, defect-free samples tend to have lower reconstruction error as they are close to the training data, while the reconstruction error of defective samples is relatively higher. However, the capacity of AE is so powerful that sometimes it can reconstruct the defective images well, especially for the defective images that share high similarity with defect-free samples.

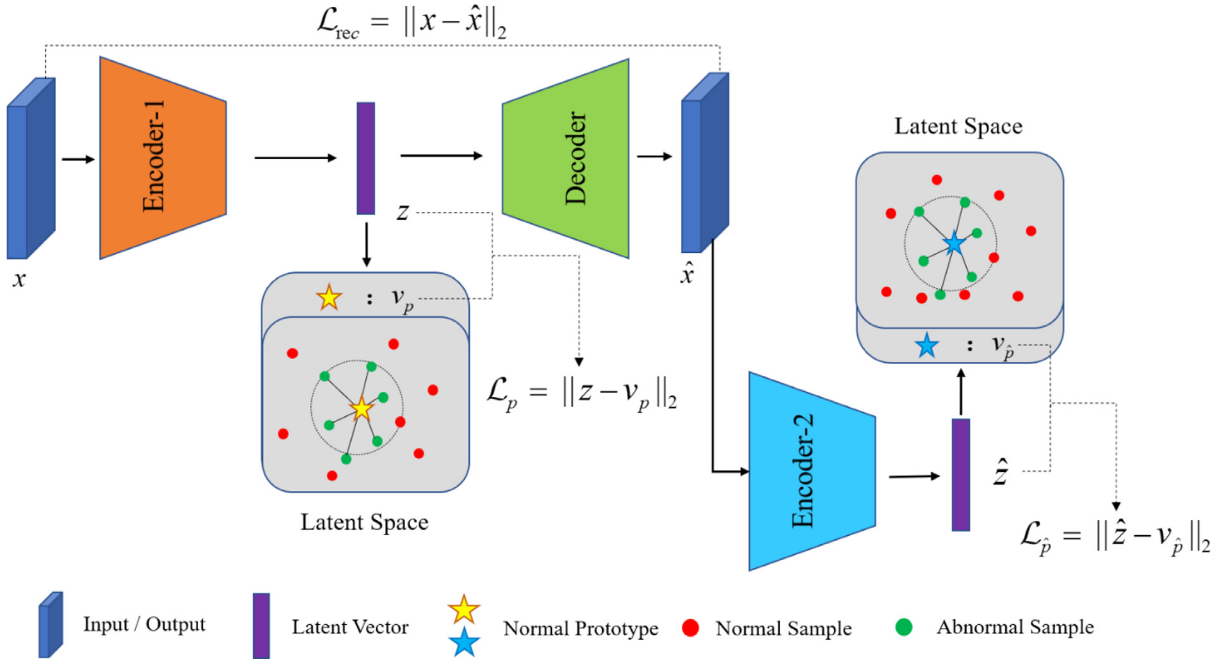


Fig. 2. The architecture of the proposed DPAE. Given the input image x, we first encode it as a latent vector z in the latent space, then the decoder is applied to get the reconstructed image $\hat{x}$, which is further encoded as another latent vector $\hat{z}$. At the training stage, the latent vector z and $\hat{z}$ of the input image are optimized to align to their prototypes v_p and $v_{\hat{p}}$, respectively.

3. Methodology

Notations. Suppose the training set $\mathrm{D} = \{x_i | i = 1, 2, \cdots, M\}$ is given, where $x_i \in \mathbf{X}$ is the ith normal sample (totally M samples) of the sample space $\mathbf{X}$. In the semi-supervised anomaly detection task, given the testing set $\mathrm{D}^* = \{(x_i^*, y_i^*) | i = 1, 2, \cdots, N\}$ from both normal and anomalous classes, where $x_i^* \in \mathbf{X}$ is the ith testing data sample (totally N samples), and $y_i^* \in \{0, 1\}$ denotes the anomaly label. Moreover, for a test samplex^*, $\mathrm{A}(x^*)$ represents the anomaly score, which can be used to determine whether $x^*$is anomaly or not. The evaluation criteria for this is to threshold $T$the score, where $\mathrm{A}(x^*) > T$ indicates anomaly.

3.1. Overview

The proposed DPAE model consists of two convolution encoders (for encoding input) and a decoder (for reconstruction). As shown in Fig. 2, given an input, we first encode it as a latent vector using encoder-1, then the decoder is applied to get the reconstructed image, which is further encoded as a latent vector by encoder-2. During training, our model is optimized to minimize the reconstruction loss between the input of the encoder-1 and the output of the decoder. At the same time, our model also minimizes the distance between the latent vectors and the normal prototypes, which are iteratively updated to represent the center of the latent vectors of defect-free samples. Given a test sample, the model with the encoder-decoder-encoder structure can automatically generate two latent vectors. As a result, the distance between the two vectors of the normal sample tends to be closer, resulting in small error for normal samples and large errors for anomalies, which can be adopted as a criterion to detect the anomalies.

3.2. Proposed model

As shown in Fig. 2, the proposed model is equipped with an encoder-decoder-encoder structure. The encoder-1 and encoder-2 are trained to map the input into an informative latent space. And the decoder is used to reconstruct the input image.

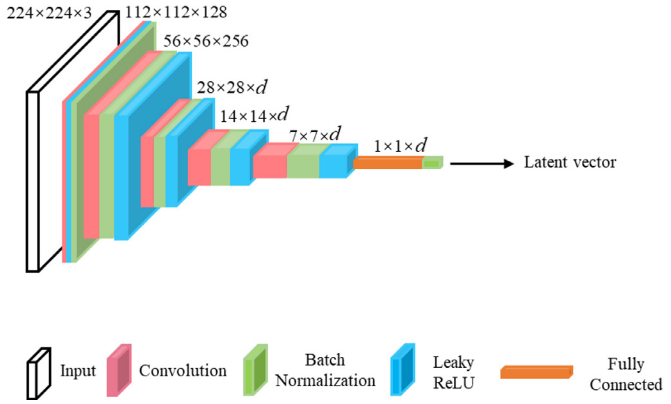


Fig. 3. The architecture of the encoder network. The input layer of shape 224 × 224 × 3 is fed into the encoder, and the output layer is the fully connected layer followed by a batch normalization layer. Finally, the output of the encoder is a latent vector with alternative size d. Best view in color.

Encoder Network. DPAE includes two encoder networks, i.e., econder-1 and encoder-2. These two encoder networks have the same structure, while the parameters of them are different. To make it concise and clear to read, we introduce the same architecture of these two encoder networks as follows. The encoder network is comprised of a sequence of blocks, including different layers: convolution, batch-normalization, leaky ReLU activation, and fully connected layer. And a visualization of our encoder network is shown in Fig. 3. With the use of the convolutional layers followed by the batch normalization and leaky ReLU activation, respectively, our encoder network is expected to down-scale the spatial resolution of feature maps as well as compress the input image to a latent vector. Notice that the dimension of the latent vector is set to d, which is tunable in our experiment.

Given an image sample $x \in \mathbf{X}$, the encoder-1 network encodes it as a latent vector $z \in Z$, which can be defined as follows:

$$z = f_e(W_e x + b_e) \quad (1)$$

Table 1
Details of the decoder network.

Stage	Type
	3 × 3 deconv, stride = 2
Block 1	Linear + BN, $C = d$
Block 2	deconv 3 × 3 + BN + Leaky ReLU, $C = d$
Block 3	deconv 3 × 3 + BN + Leaky ReLU, $C = d$
Block 4	deconv 3 × 3 + BN + Leaky ReLU, $C = 256$
Block 5	deconv 3 × 3 + BN + Leaky ReLU, $C = 128$
Block 6	deconv 3 × 3 + BN + Leaky ReLU, $C = 3$

where W_e and b_e is the weight and bias of encoder-1, respectively, and f_e is the nonlinear transformation function.

Decoder Network. The decoder network usually cooperates with the encoder network to reconstruct the input image from the latent vector z. And the details of our decoder network are presented in Table 1. Firstly, a linear layer followed by batch normalization is applied to upscale the latent vector z. Then, five blocks comprised by deconvolutional transpose layers, batch normalization, and leaky ReLU activation are utilized to reconstruct the original image$x$ as $\hat{x}$. Therefore, the decoding process can be formulated as follows:

$$\hat{x} = f_d(W_d z + b_d) \tag{2}$$

where W_d and b_d is the weight and bias of the decoder network, respectively, f_d is the nonlinear transformation function, and $\hat{x}$ is the reconstructed form of the input image x.

Encoder-Decoder-Encoder Structure. Unlike previous autoencoder-based approaches, an encoder-decoder-encoder structure is adopted in the proposed model. Besides the encoder-1 and decoder, encoder-2 is also leveraged to downscale the reconstructed image $\hat{x}$ by compressing it to a latent vector $\hat{z}$. With different network parameters, encoder-2 has the same architecture as encoder-1. Finally, the latent vector $\hat{z}$ can be described as:

$$\hat{z} = f_e(\hat{W}_e \hat{x} + \hat{b}_e) \tag{3}$$

where $\hat{W}_e$ and $\hat{b}_e$ is the weight and bias of encoder-2, respectively.

3.3. Model training

Adopting autoencoder structure, anomaly detection models are expected to minimize the reconstruction error on normal images during training and induce large reconstruction error on anomalies at the test stage. However, because of the diversity of the normal images and high similarity between normal and anomalous images, the reconstruction error is not effective enough to distinguish normal and anomalous images.

Therefore, encoder-2 is used to map the reconstructed images $\hat{x}$ to a latent vector $\hat{z}$. During training, z and $\hat{z}$ are optimized to align to their prototypes, respectively, the dissimilarity between z and $\hat{z}$ are minimized at the same time. For anomalous images, the dissimilarity between the latent vectors z and $\hat{z}$ will be enlarged. As a result, given a test image x^*, high dissimilarity between the latent vectors z and $\hat{z}$ indicates anomaly. Therefore, three loss functions are combined in our objective function, and it is proved to be effective for the image anomaly detection task.

Reconstruction Loss. Given the training set $\mathrm{D} = \{x_i | i = 1, 2, \cdots, M\}$ containing M samples, we firstly consider the distance between the input image x and its reconstruction $\hat{x}$. The reconstruction error on each sample is minimized as follows:

$$\mathrm{L}_{rec}(x, \hat{x}) = \ || \ x - \hat{x} \ ||_2^2 \tag{4}$$

where the ℓ_2-norm is used to measure the reconstruction error.

Dual Prototype Loss. Due to the great capacity of the convolutional neural network (CNN), some anomalous images can even be effectively reconstructed, resulting in an ambiguous decision boundary of normal and anomaly. Therefore, minimizing the intra-class variations of normal samples while keeping the features of normal and anomaly separable is key to the anomaly detection task. To this end, the latent vector of the normal sample is restricted to keep close to the normal prototype by the following loss function:

$$\mathrm{L}_p = \frac{1}{2} \sum_{i=1}^{m} || \ z_i - \upsilon_p ||_2^2 \tag{5}$$

where z_i is the first latent vector of the i th training sample, $\nu_p \in \mathbf{X}$ is the prototype of z_i, and m is the size of mini-batch. This formulation can effectively narrow the intra-class distance between the normal samples. For the normal prototype ν_p, it is initialized as random vector with the same size as z_i and updated as the deep features changed. Instead of updating ν_p based on the entire training set, we implement update with respect to the mini-batch. And the gradient of L_p with respect to z_i is computed as:

$$\frac{\partial \mathrm{L}_p}{\partial z_i} = z_i - \upsilon_p \tag{6}$$

To update the normal prototype, the update equation is formulated as follows:

$$\Delta \upsilon_p = \frac{1}{m} \sum_{i=1}^{m} (\upsilon_p - z_i) \tag{7}$$

Besides the prototype loss of z_i, we restrict the latent vector $\hat{z}_i$ to enclose to its prototype, and the prototype loss of $\hat{z}_i$ is defined as:

$$\mathrm{L}_{\hat{p}} = \frac{1}{2} \sum_{i=1}^{m} ||\hat{z}_i - \upsilon_{\hat{p}}||_2^2 \tag{8}$$

where $\hat{z}_i$ is the second latent vector of the normal samples x_i, and $\upsilon_{\hat{p}} \in \mathbf{Z}$ is the normal prototype of $\hat{z}_i$.

Overall, the final loss function of our DPAE model is formed as a combination of the reconstruction loss and dual prototype loss:

$$\mathrm{L} = \sum_{i=1}^{m} ||x_i - \ \hat{x}_i||_2^2 + \frac{\lambda}{2} \sum_{i=1}^{m} ||z_i - \upsilon_p||_2^2 + \frac{\gamma}{2} \sum_{i=1}^{m} ||\hat{z}_i - \upsilon_{\hat{p}}||_2^2 \tag{9}$$

where λ and γ is the weight coefficients to regulate the importance of different losses.

3.4. Anomaly assessment

Anomaly Score. After training the proposed model on the normal samples, instead of using reconstruction error to determine whether a given image is normal or not, we suggest to predict anomalies by evaluating the mean squared error (MSE) between z and $\hat{z}$. Therefore, given a test sample x^*, the anomaly score is defined as:

$$\mathrm{A}(x^*) = ||z - \hat{z}||_2^2 \tag{10}$$

where z and $\hat{z}$ are the outputs of encoder-1 and encoder-2, respectively, i. e., the latent vectors. Furthermore, the set of anomaly scores is defined as $\mathrm{S} = \{s_i : \mathrm{A}(x^*), x^* \in \mathrm{D}^*\}$, and then we normalize the anomaly scores to the range of [0,1] by the Eq. (11).

$$\hat{s}_i = \frac{s_i - \min(\mathrm{S})}{\max(\mathrm{S}) - \min(\mathrm{S})} \tag{11}$$

Finally, given a test image x^*, $\hat{s}_i$ is utilized as the anomaly score to distinguish normal and anomaly. The value of $\hat{s}_i$ closer to 1 indicates the image is more likely a defective image.

Evaluation Metric. We mainly choose the AUC (Area Under Curve) as the criterion for performance evaluation, and it is obtained by calculating the area under the Receiver Operation Characteristic (ROC) with varying thresholds. In addition, following the standard protocol [37], we also consider the average precision, sensitivity (TPR, i. e., true positive rate), specificity (TNR, i. e., true negative rate), and F1 score after 20 runs as an intuitive way to compare the anomaly detection performance. Specifically, based on the defect ratio listed in Table 2, we

Table 2
Statistical overview of the four surface defect datasets.

DATASET	Train	Test defect-free	Test defective	Total images	Defect groups	Defect ratio
APSD	500	150	450	1100	11	0.75
RSD	500	150	450	1100	1	0.75
MT	330	142	422	894	5	0.75
CSD	280	28	89	397	5	0.76

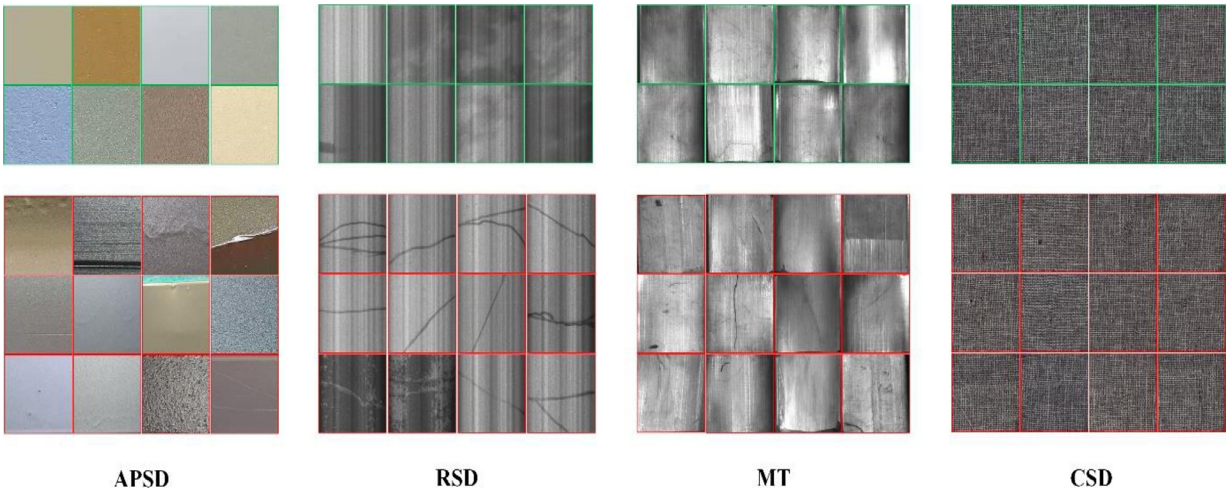


Fig. 4. Examples of the real images of four surface defect datasets. Top rows: defect-free images. Bottom rows: defective images.

chose the threshold to distinguish defective samples. For example, when DPAE performs on the APSD dataset, the top 70% samples of the highest anomaly scores will be marked as anomalies.

4. Experimental setup

4.1. Datasets

To evaluate our anomaly detection framework, we first construct the Aluminum Profile Surface Defect (APSD) dataset, and the raw data come from a defect-recognition competition [15]. The raw data are collected manually with non-uniform sampling conditions, such as different light and focal length. Therefore, original images are processed to adapt to the anomaly detection task. Besides, to evaluate the effectiveness and generalization of the anomaly detection methods, our experiments are also conducted on three other datasets, i. e., the Road Surface Defect (RSD) dataset, the Magnetic-Tile defect (MT) dataset [14], and the Carpet Surface Defect (CSD) dataset [3]. Exemplary images for anomalous and normal classes for all four datasets are shown in Fig. 4. And the statistical details of the five surface defect datasets are listed in Table 2.

APSD. This dataset comprises ten categories surface defects from the aluminum profile, including blister, bump, coarseness, coating crack, convexity, damage, dirty spot, indentation, jet flow, orange peel, and rub mark. And there are 650 defect-free images and 450 defective images with a resolution of 224 × 224. In the experiment, 500 defect-free (normal) images serve as training samples, and 150 defect-free and 450 defective (anomalous) images are utilized to test the effectiveness of our method. By the way, defect ratio represents the percentage of defective samples in the test set.

RSD. The RSD dataset contains tarmac images with two classes, namely positive and negative inlaid patch. 800 inlaid patch images with the size of around 3000 × 2000 pixels are collected by [7]. In this experiment, 450 inlaid patch images are randomly resized to 224 × 224 to strengthen the diversity of the defects. Besides, 650 defect-free images are cropped and resized to 224 × 224 by us to obtain normal samples. Similar to the APSD dataset, 500 defect-free images are required in the training stage, and the test set contains 150 defect-free and 450 defective images.

MT. The magnetic-tile dataset collected by [14] consists of 1344 images of the magnetic tile surface. And the MT dataset includes 422 defective images from 5 groups, i. e., blowhole, break, crack, fray, and uneven, and 472 defect-free images, all these images have different resolutions. Most of these images contain a series of noise, e. g., complex texture, and different illumination intensity, bringing a great challenge for anomaly detection algorithm. In our anomaly detection experiment, 330 defect-free images are used as training samples, and 142 defect-free and 422 defective images serve as test samples.

CSD. Carpet surface defect dataset [3] contains 308 defect-free and 89 defective carpet images from 5 types of defects. Most of the defective regions in the CSD dataset are very small, with low-contrast, and share high similarity with defect-free images. Followed by the setting in [3], 280 defect-free images are used in the training stage, 28 defect-free and 89 defective images serve as test samples to evaluate the anomaly detection model.

4.2. Implementation details

Our approach is implemented in PyTorch (1.2.0 with python 3.7) by optimizing the networks using the SGD optimizer with an initial learning rate $lr = 0.001$, weight decay 0.0005, and momentum 0.9. Besides, all the training images are reshaped into 224 × 224 × 3 and randomly flipped to improve the diversity of the training samples. The proposed method is trained to minimize the weighted loss defined in Eq. (9) using fixed weight values $\lambda = 1$ and $\beta = 1$. And we train our model for 50 epochs with batch size 8 on all the four datasets. Also, all our experiments are executed on a PC with an Intel(R) Core (TM) i7–7700 3/4 GHz processor, 8GB RAM, and an 8 GB Nvidia 1070Ti Xp. The source code and datasets are available at https://github.com/JasonLiu-Dr/DPAE.

5. Experimental results

In this section, we validate the proposed DPAE for anomaly detection. Our experiments are conducted on four datasets and compared with five anomaly methods, i. e., convolutional autoencoder with L_2 loss (AE_{L2}) [40], convolutional autoencoder with SSIM loss (AE_{SSIM}) [40], variation autoencoder [33], AnoGAN [11], and GANomaly [10]. In the APSD dataset, we evaluate the performance of anomaly detection methods in terms of five aforementioned evaluation criteria, which are precision, sensitivity, specificity, F1 score, and AUC, respectively. In addition, AUC and F1 score are used to verify the generalization performance on the other datasets, namely, RSD, MT, and CSD.

5.1. Results on the APSD dataset

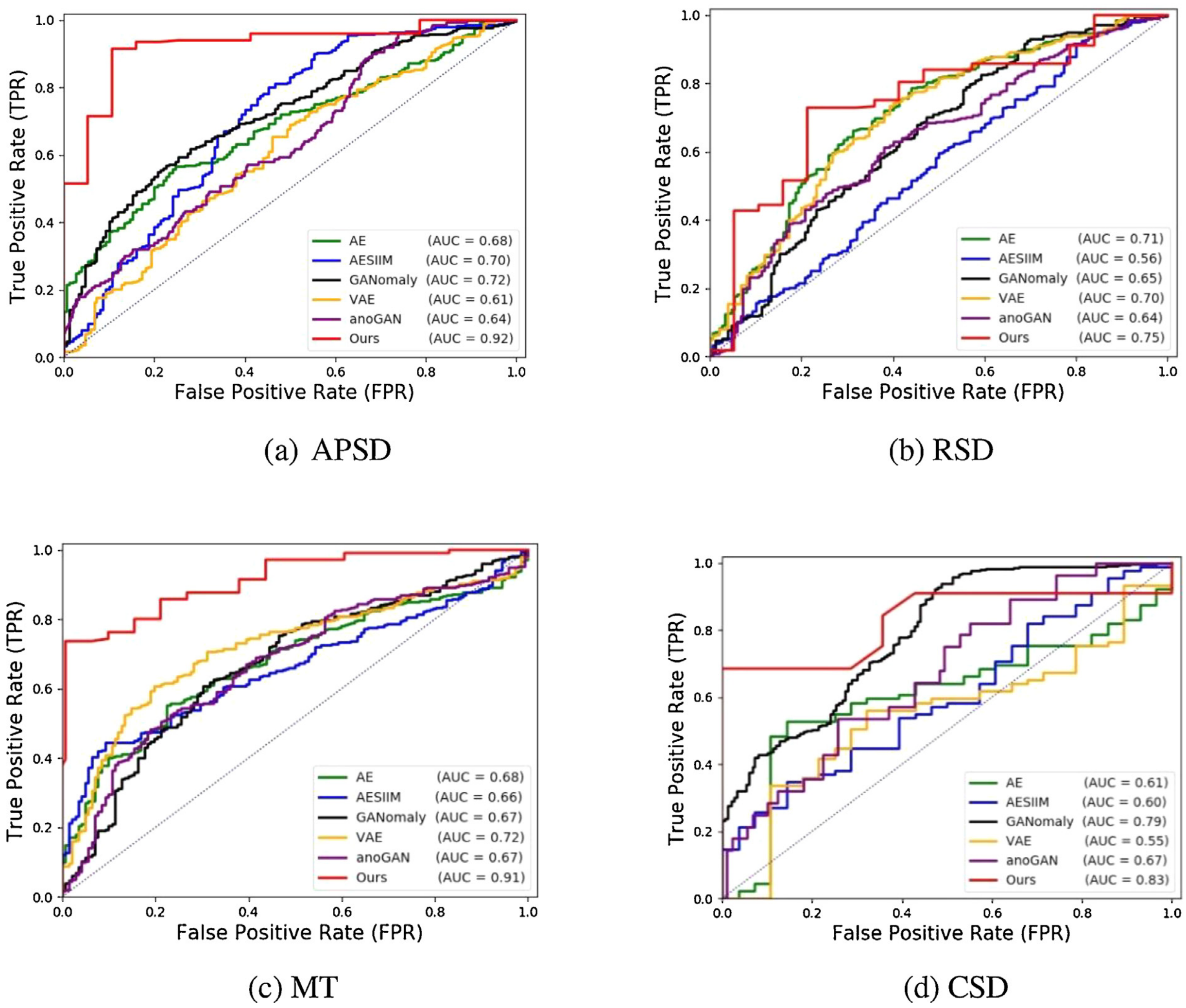
As mentioned above, the low availability of defective samples is an inherent problem in the surface inspection tasks, which leads to the limitation of the application of supervised learning methods. Therefore, our DPAE model aims to accomplish the anomaly detection task merely using defect-free samples. To this end, we conduct comprehensive experiments on the APSD dataset, and the results are summarized in Table 3.

Compared with the rest of the approaches, DPAE achieves the best performance in terms of all the evaluation criteria. Specifically, DPAE obtains a 13% improvement than the competitive baseline AE_{SSIM} on accuracy. And DPAE has a sensitivity of 0.92 and specificity of 0.89, which indicates that both the normal and anomalous images can be well distinguished. Besides, DPAE outperforms other state-of-the-art methods in F1-socre, which further demonstrates that our model can accurately identify not only anomalous images but also normal images. Furthermore, our model exceeds all other recent methods with an improvement ranging from 20% to 29% in AUC with different threshold values.

A pivotal aspect of the anomaly detection system is the detection robustness, i. e., performance when varying the classification threshold. Fig. 5(a) presents the Receiver Operating Characteristic (ROC) curve of six methods on the APSD dataset. As can be viewed, compared with other five methods, our model exhibits high robustness to classification threshold variation.

5.2. Generalization performance

To validate the generalization performance of the DPAE model, we further implement experiments on the other three datasets, i. e., RSD, MT, and CSD, and the results are presented in Table 4. Note that only few parameters need to be tuned when our method is applied on different datasets. We see that the AUC and F1 score of all the evaluated methods on the RSD dataset are relatively poor than that on the other three datasets. In fact, this phenomenon can be explained by the slender shape of road defects. Convolutional neural network (CNN) usually exhibits inferior performance when recognizing tiny objects. Compared with defects in other shapes, after several convolutional and deconvolutional operations, slender road defects tend to lose more information. As a result, for the latent vectors of slender defects, very little information could be saved and used to distinguish anomaly.



(a) APSD

(b) RSD

(c) MT

(d) CSD

Fig. 5. Receiver operating characteristic (ROC) curves and corresponding area under the ROC curve (AUC) values (specified in parentheses) of different anomaly detection methods on four defect datasets (APSD, CSD, MT, and RSD).

Table 3
Experimental results of different methods on the APSD dataset. The best results are marked in bold, and the second best are unberlined.

DATASET	Precision	Sensitivity (TPR)	Specificity (TNR)	F1-score	AUC
AE_{L2} [40]	0.78	0.74	0.45	0.76	0.68
AE_{SSIM} [40]	0.83	0.64	0.66	0.73	0.67
VAE [33]	0.79	0.64	0.54	0.71	0.63
AnoGAN [11]	0.69	0.68	0.58	0.63	0.64
GANomaly [10]	0.79	0.74	0.49	0.76	0.72
DPAE (Ours)	**0.96**	**0.92**	**0.89**	**0.94**	**0.92**

Table 4
Experiment results of different methods on the four datasets. The best results are marked in bold, and the second best are underlined.

METHOD	APSD		RSD		MT		CSD	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
$AE(L_2)$	0.68	0.76	0.71	0.77	0.68	0.73	0.61	0.73
AE(SSIM)	0.70	0.73	0.56	0.62	0.66	0.65	0.60	0.76
VAE [33]	0.61	0.71	0.70	**0.82**	0.72	**0.78**	0.55	0.69
AnoGAN [11]	0.64	0.63	0.64	0.79	0.67	0.73	0.67	0.76
GANomaly [10]	0.72	0.76	0.65	0.66	0.67	0.70	0.79	0.64
DPAE (Ours)	**0.92**	**0.94**	**0.75**	0.69	**0.91**	**0.78**	**0.83**	**0.85**

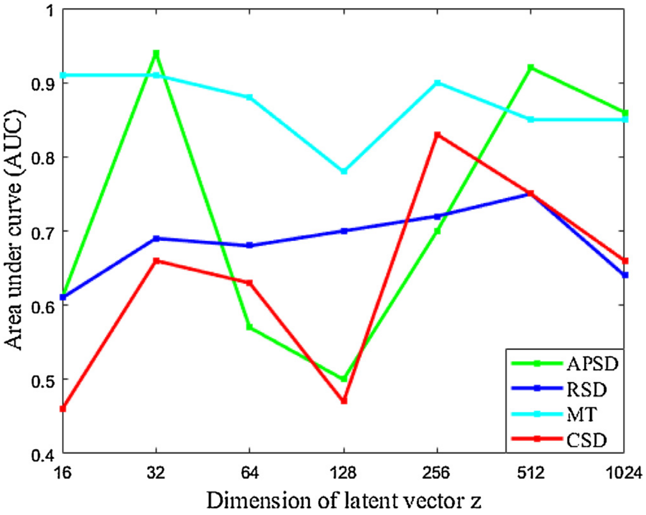


Fig. 6. Overall performance of our model based on varying dimension of the latent vector *z* on different datasets.

For the MT dataset, DPAE beats all the five methods with an improvement ranging from 19% to 25% in AUC, and achieves a competitive F1 score of 0.78, which is the same as VAE. Besides, for the more challenging datasets CSD, DPAE obtains 19% and 13% average improvement than other state-of-the-art methods in AUC and F1 score, respectively.

Fig. 5(b), (c), and (d) show the ROC curves of different methods on the datasets RSD, MT, and CSD, respectively. Compared with the previous models, DPAE achieves state-of-the-art performance over all the three datasets, especially on the MT dataset. And the experiments demonstrate that the proposed method is effective and robust for the anomaly detection task on different kinds of defect datasets.

5.3. Ablation study

To show the effectiveness of our method, some ablative settings are designed. And the APSD dataset is taken as an example for ablation analysis. We study three aspects that influence the performance of DPAE in the anomaly detection task, i. e., latent vector size, loss function, and anomaly score.

Table 5
Ablation studies based on the APSD dataset. The best result is marked in bold, and the second best is underlined.

Method	AUC
$AE(L_2)$	0.68
AE(SSIM)	0.67
DPAE w/o RS	0.72
DPAE w/o DPS	0.54
DPAE	**0.92**

Latent vector size in DPAE. We present the trend of AUC w.r.t. varying values of latent vector size *d*over a range of {16, 32, 64, 128, 256, 512, 1024}. Anomaly detection results are shown in Fig. 6. We can see that DPAE with small values of *d* achieves better AUC results on the APSD and MT dataset. In particular, the AUC of the proposed method on the APSD and MT dataset when $d = 32$ are 0.94 and 0.89, respectively. Besides, DPAE achieves better AUC results on the RSD and CSD dataset when *d*is a large value. Specifically, DPAE obtains the AUC of 0.75 on the RSD dataset when $d = 512$, and 0.83 on the CSD dataset when $d = 256$.

Loss function in DPAE. We also performed an ablation study on the APSD dataset to illustrate the effectiveness of each loss function of DPAE. And five scenarios are considered: autoencoder with L2 loss (AE(L2)), autoencoder with SSIM loss (AE(SSIM)), DPAE with only reconstruction loss (RS), DPAE with only dual prototype loss (DPS), and DPAE with both RS and DPS. The results are presented in Table 5. We can see that DPAE with only RS achieves an AUC of 0.72, which is slightly higher than that of AE(L2) and AE(SSIM). The AUC of DPAE with DPS is much lower than other methods, and this phenomenon indicates that the anomaly detection methods with only dual prototype loss are difficult to distinguish the normal and anomaly. Finally, our model with both RS and DPS achieves the best performance, which illustrates the effectiveness of the combination of reconstruction loss and dual prototype loss in the training stage of the anomaly detection task.

Anomaly score in DPAE. Anomaly score is the crucial component that directly influences the performance of anomaly detection methods. And four types of anomaly scores are considered in this section: RLX, i. e., reconstruction loss in Eq. (4); RLZ, i. e., reconstruction loss in Eq. (10); PL1, the mean square loss between the first normal prototype and the first latent vectors of the test samples; PL2, the mean square loss

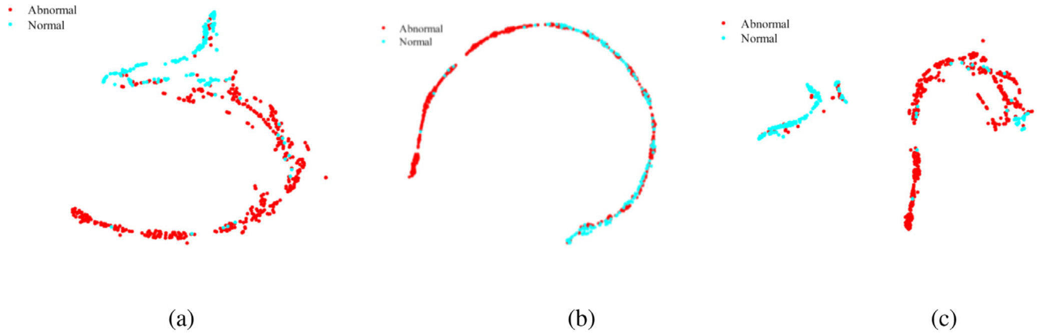


Fig. 7. T-SNE visualization of test samples on the APSD dataset. (a) Latent vectors of encoder-1. (b) Latent vectors of encoder-2. (c) The vectors comprised of the difference of the first and second latent vectors.

Table 6
Effectiveness of different types of anomaly score. The best result is marked in bold, and the second best is underlined.

Anomaly score	AUC
RLX	0.73
RLZ	**0.92**
PL1	0.90
PL2	0.78

between the second normal prototype and the second latent vectors of the test samples.

We present the AUC results of DPAE with the aforementioned anomaly scores on the APSD dataset. The results are shown in Table 6. We can see that our model with RLZ obtains an AUC of 0.92, which is 19% higher than our model with RLX. Besides, DPAE with PL1 achieves an AUC of 0.90, whereas the AUC of DPAE with PL2 is 0.78. Overall, the results demonstrate the effectiveness of taking RLZ as anomaly score and the potential of PL1 as anomaly score. Besides, we visualize different features of the test samples, and the results are presented in Fig. 7. We can see that the distribution of the normal and anomalous samples in (c) is more separable, which illustrates the effectiveness of using the anomaly score in Eq. (10).

6. Conclusion

In this paper, we proposed a novel dual prototype auto-encoder (DPAE) for surface defect datasets based on semi-supervised anomaly detection methods. The proposed DPAE follows an encoder-decoder-encoder paradigm, and it is trained with the guidance of the combination of reconstruction loss and dual prototype loss. Therefore, the latent vectors of inputs are restricted to align to their prototype, and large distance between latent vectors can be used as an indicator of anomalies. Besides, we constructed the APSD dataset for the anomaly detection task. Extensive experiments conducted on the APSD dataset and the other three defect datasets prove the generalization and effectiveness of the proposed method. Future work will consider pixel-wise anomaly detection of surface images of the industrial products.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

We wish to draw the attention of the Editor to the following facts which may be considered as potential conflicts of interest and to significant financial contributions to this work.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (51805078, 51374063), the National Key Research and Development Program of China (2017YFB0304200), Taizhou Science and Technology Planning Project (1902GY09).

References

[1] He Y, Song KC, Meng QG, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Trans Instrument Measur 2020;69(4):1493–504.

[2] Gong D, Liu L, Le V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection. In: Proceedings of the International Conference on Computer Vision (ICCV); 2019. p. 1705–14.

[3] Bergmann P, Fauser M, Sattlegger D, et al. MVTec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 9592–600.

[4] Zhang D, Song KC, Xu J, et al. Unified detection method of aluminum profile surface defects: common and rare defect categories. Opt Lasers Eng 2020;126:105936.

[5] Fu G, Sun P, Zhu W, et al. A deep-learning-based approach for fast and robust steel surface defects classification. Opt Lasers Eng 2019;121:397–405.

[6] Zhuang B, Shen C, Tan M, et al. Structured binary neural networks for accurate image classification and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 413–22.

[7] Dong HW, Song KC, He Y, et al. PGA-Net: pyramid feature fusion and global context attention network for automated surface defect detection. IEEE Trans Ind Infor 2019. doi:10.1109/TII.2019.2958826.

[8] Zimek A, Schubert E, Kriegel H P. A survey on unsupervised outlier detection in high-dimensional numerical data. Stat Anal Data Mining: ASA Data Sci J 2012;5(5):363–87.

[9] Xiong L, Póczos B, Schneider J G. Group anomaly detection using flexible genre models. In: Advances in Neural Information Processing Systems (NIPS); 2011. p. 1071–9.

[10] Akcay S, Atapour-Abarghouei A, Breckon T P. GANomaly: semi-supervised anomaly detection via adversarial training. In: Asian Conference on Computer Vision; 2018. p. 622–37.

[11] Schlegl T, Seeböck P, Waldstein SM, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging; 2017. p. 146–57.

[12] Chen Y, Zhou XS, Huang TS, et al. One-class svm for learning in image retrieval. In: Proceedings of the International Conference on Image Processing (ICIP); 2001. p. 34–7.

[13] Zimmerer D, Isensee F, Petersen J, et al. Unsupervised anomaly localization using variational auto-encoders. In: International Conference on Medical Image Computing and Computer-assisted Intervention; 2019. p. 289–97.

[14] Huang Y, Qiu C, Yuan K. Surface defect saliency of magnetic tile. In: International Conference on Automation Science and Engineering (CASE); 2018. p. 612–17.

[15] https://tianchi.aliyun.com/competition/entrance/231682/information; 2018.

[16] He Y, Song KC, Dong HW, et al. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Opt Lasers Eng 2019;122:294–302.

[17] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv 2009;41(3):1–58.

[18] He D, Xu K, Zhou P, et al. Surface defect classification of steels with a new semi-supervised learning method. Opt Lasers Eng 2019;117:40–8.

[19] Polzleitner W. Defect detection on wooden surface using Gabor filters with evolutionary algorithm design. In: International Joint Conference on Neural Networks (IJCNN); 2001. p. 750–5.

[20] Mak KL, Peng P, Yiu K, et al. Fabric defect detection using morphological filters. Image Vis Comput 2009;27(10):1585–92.

[21] Chen S, Feng J, Zou L. Study of fabric defects detection through Gabor filter based on scale transformation. In: International Conference on Image Analysis and Signal Processing; 2010. p. 97–9.

[22] Munkhdalai L, Munkhdalai T, Ryu KH, et al. GEV-NN: a deep neural network architecture for class imbalance problem in binary classification. Knowledge-Based System 2020. doi:10.1016/j.knosys.2020.105534.

[23] Parzen E. On estimation of a probability density function and mode. The annals of mathematical statistics 1962;33(3):1065–76.

[24] Bezdek JC, Ehrlich R, Full W. Fcm: the fuzzy c-means clustering algorithm. Computers and Geosciences 1984;10(2):191–203.

[25] Campbell C, Bennett PB. A linear programming approach to novelty detection. In: Advances in Neural Information Processing Systems (NIPS); 2001. p. 395–401.

[26] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[27] Racki D, Tomazevic D, Skocaj D. A compact convolutional neural network for textured surface anomaly detection. In: IEEE Winter Conference on Application of Computer Vsion (WACV); 2018. p. 1331–9.

[28] Zhou JT, Du J, Zhu H, et al. AnomalyNet: an anomaly detection network for video surveillance. IEEE Trans Infor Forensics Secur 2019;14(10):2537–50.

[29] Ergen T., Mirza A.H., Kozat S.S., et al. Unsupervised and semi-supervised anomaly Detection with LSTM neural networks. 2017. arXiv:1710.09207.

[30] Minhas M.S., Zelek J. Anomaly Detection in Images. 2019. arXiv:1905.13147.

[31] Star B, Lutjen M, Freitag M. Anomaly detection with convolutional neural networks for industrial surface inspection. In: Conference on Intelligent Computation in Manufacturing Engineering (CIRP); 2018. p. 484–9.

[32] Awoyemi JO, Adetunmbi AO, Oluuwadare SA. Credit card fraud detection using machine learning techniques: a comparative analysis. In: International Conference on Computing Networking and Informatics (ICCNI); 2017. p. 1–9.

[33] Kingma DP, Welling M. Auto-encoding variational Bayes. International Conference on Learning Representations; 2014.

[34] Zhao Y., Ding X., Yang J., et al. SUOD: toward Scalable Unsupervised Outlier Detection. 2020. arXiv:2002.03222.

[35] Huang X, Luo X, Wang R. A real-time parallel combination segmentation method for aluminum surface defect images. In: International Conference on Machine Learning and Cybernetic (ICMLC); 2015. p. 544–9.

[36] Xu K, Liu SH, Ai YH. Application of Shearlet transform to classification of surface defects for metals. Image Vis Comput 2015;35(3):23–30.

[37] Zong Bo, Song Qi, Min Martin Renqiang. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. International Conference on Learning Representations (ICLR); 2018.

[38] Chalapathy R., Chawla S. Deep learning for anomaly detection: a survey. arXiv preprint arXiv:1901.03407; 2019.

[39] Ruff L, Vandermeulen RA, Gornitz N. Deep semi-supervised anomaly detection. International Conference on Learning Representation (ICLR); 2020.

[40] Bergmann P, Löwe S, Fauser M, Sattlegger D, Steger C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In: International Conference on Computer Vision Theory and Applications (VISAPP); 2019. p. 372–80.

[41] Mastan I D, Raman S. Multi-level encoder-decoder architectures for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 1728–37.

[42] Niu Menghui, Song Kechen, Huang Liming, Wang Qi, Yan Yunhui, Meng Qinggang. Unsupervised Saliency Detection of Rail Surface Defects using Stereoscopic Images. IEEE Transactions on Industrial Informatics 2020. doi:10.1109/TII.2020.3004397.