SURVEY

A Survey on Unsupervised Anomaly Detection Algorithms for Industrial Images

YAJIE CUI^{ID}, ZHAOXIANG LIU^{ID}, AND SHIGUO LIAN, (Member, IEEE)
Unicom Digital Technology, China Unicom, Beijing 100013, China
Corresponding authors: Zhaoxiang Liu (liuzx178@chinaunicom.cn) and Shiguo Lian (liansg@chinaunicom.cn)

ABSTRACT In line with the development of Industry 4.0, surface defect detection/anomaly detection becomes a topical subject in the industry field. Improving efficiency as well as saving labor costs has steadily become a matter of great concern in practice, where deep learning-based algorithms perform better than traditional vision inspection methods in recent years. While existing deep learning-based algorithms are biased towards supervised learning, which not only necessitates a huge amount of labeled data and human labor, but also brings about inefficiency and limitations. In contrast, recent research shows that unsupervised learning has great potential in tackling the above disadvantages for visual industrial anomaly detection. In this survey, we summarize current challenges and provide a thorough overview of recently proposed unsupervised algorithms for visual industrial anomaly detection covering five categories, whose innovation points and frameworks are described in detail. Meanwhile, publicly available datasets for industrial anomaly detection are introduced. By comparing different classes of methods, the advantages and disadvantages of anomaly detection algorithms are summarized. Based on the current research framework, we point out the core issue that remains to be resolved and provide further improvement directions. Meanwhile, based on the latest technological trends, we offer insights into future research directions. It is expected to assist both the research community and industry in developing a broader and cross-domain perspective.

INDEX TERMS Industrial anomaly detection, unsupervised learning, deep learning.

I. INTRODUCTION

Industry 4.0 is an era of making use of information technology to promote the industrial revolution, that is, the intelligent era. It is the fourth industrial revolution dominated by intelligent manufacturing. Adhere to the development trend of Industry 4.0, it is the general trend to build a smart manufacture system.

Ideally, once a production link deviates from the standard operation, an alarm signal will be sent, and the producer can make positive improvement response in the shortest time. This transparent and efficient information based production process can minimize production costs and also avoid wasting materials. In the long run, smart manufacturing mode based on artificial intelligence technology [1] can reduce the requirements on human decision, such as dependence on

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy^{ID}.

technical experts, by mining and depositing relevant knowledge, so that labor costs can be saved.

Material anomaly extensively exists in industrial production [1], [2], [3], and more and more safety problems are caused by material defects. Hence people pay more attention to the detection of material anomalies. There are a lot of defects images in the industrial scene. Examples of surface defects are shown in Fig. 1. The traditional method of surface anomaly detection and localization relies on the manual operation by qualified specialists, which is not only inefficient but also depends on the subjective judgment of operators, making it difficult to ensure the accuracy of detection. The production mode of anomaly detection equipment combined with industrial production line ensures the quality of products, reduces the cost of manual testing, and improves the efficiency of production as well. With the rapid rise of computer image processing technology, many algorithms have been gradually applied to the field of material anomaly detection, improving the accuracy of material anomaly detection. In recent

years, deep learning [4], [5] has also been applied to material anomaly detection and achieved extraordinary performance.

Current deep learning-based supervised algorithms have certain limitations. Model training requires a large amount of labeled data [5], while images with defect labels are not easy to obtain. And the lack of defect samples makes it difficult to bring the models online, which also limits the application of deep learning in the industrial detection field. Therefore, a new solution is urgently needed - unsupervised algorithms, which need no labeled data. This paper provides a review of some recently proposed unsupervised methods, whose innovation points and frameworks are described in detail. Note that we only concentrate on industrial vision anomaly detection algorithms. Particularly, the industry uses the terms defect detection, visual anomaly detection, and surface detection, which we all count in our research. Meanwhile, publicly available datasets for industrial anomaly detection are introduced, with experimental results displayed. This review offers different contributions that distinguish it from other reviews.

- We discuss the inadequacies of the current algorithms and dataset tailored to the requirements of actual industrial scenarios, such as edible oil impurity detection.
- Based on the current research framework, we suggest that the conflict between FAR (false alarm rate) and MAR (missed alarm rate) is the core issue that remains to be resolved. We also provide further improvement directions to current methods, like integration of diverse technologies.
- We offer insights into future directions based on the latest technological trends, such as foundation model and multimodal learning.

The subsequent content of the article is organized as follows: related works in Section II, visual anomaly detection methodology in Section III, comparison and analysis of methodologies in Section IV, introduction to industrial datasets in Section V, challenge and discussion of research actuality and future development direction in Section VI, conclusion in Section VII.

II. RELATED WORKS
A. PREVIOUS ALGORITHMS
We only focus on the surprising success and dominance of anomaly detection in industrial images but exclude other areas such as action recognition [6], [7] and video anomaly detection [8]. Although some of the strategies have been validated in the above scenario, real industrial images lack prior knowledge of action images and video sequence information, which makes it difficult for models to generalize across different domains.

For some existing methods of visual anomaly detection, the development and changes in technology are introduced below. Initially, anomaly detection relied on statistical meth-

ods. Statistical approaches assess the geographic distribution of pixel values by extracting statistical information from defect images. Histogram information [9], [10], [11], [12], [13], [14], co-occurrence matrices [15], [16], [17], [18], [19], [20], and local binary patterns (LBP) [21], [22], [23], [24], [25], [26] have all been presented as statistical methods for defect detection. Statistical approaches can present anomalies in an intuitive and discriminative manner, and they are simple to model, interpret, and display. However, they frequently make assumptions, such as separable defect regions, that cannot be satisfied in all scenarios. Later, hand-extracted features can describe the structure of the image. In structural methods, defect feature is characterized by texture elements [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37]. As a result, the structural approaches' goals are to extract the texture elements of defects, which are used to represent the spatial placement rules. The geometrical feature can be found using structural approaches. This approach is easier to implement and better suited to random textured defects. However, the majority of them are sensitive to the shape and size of defects, and defect images should be aperiodic.

In the field of image processing, filter-dependent methods are also used for anomaly detection. Filter-based methods apply some filter banks on defect images and calculate the energy of the filter responses [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50]. Common filter-based methods include Sobel operator, Canny operator, Gabor operator, Laplacian operator, wavelet transform, and Fourier transform, which can be further divided into spatial domain, frequency domain, and spatial-frequency domain methods. In vision-based anomaly recognition, filter-based approaches are widespread. The cross-domain methods can aid the model in extracting more meaningful information. Furthermore, they are affine transformation invariant and can handle multi-scale defects. While, they may not be appropriate to random textured images, and some of them may be influenced by feature correlations and noises.

With the development of neural networks and machine learning, a large number of supervised algorithms have appeared. Supervised Neural Networks [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], Support Vector Machines (SVM) [63], [64], [65] and k-Nearest Neighbors [66], [67], [68], [69], [70], [71], [72] are the most common supervised algorithms. Recently, deep learning-based algorithms are becoming popular. The majority of deep learning-based visual anomaly detection is data-driven. To build the visual anomaly detection model, supervised methods take two means. The first one trains an image-level classification model, which requires a labeled training set including both normal and abnormal samples. The second conducts refined object localization. Containing more information, supervised methods should theoretically yield higher detection rates than semi-supervised and unsupervised methods. However, because of lacking training dataset that covers all defect locations and erroneous labeling, there still exists certain technical difficulties.

---

[1]http://faculty.neu.edu.cn/yunhyan/NEU_surface_defect_database.html

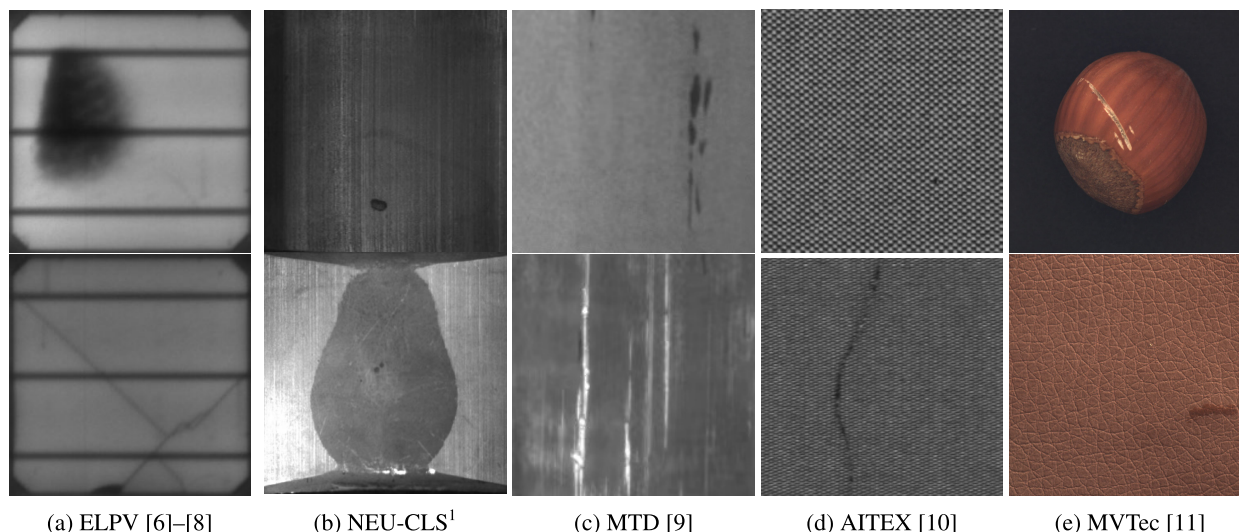| (a) ELPV [6]–[8] | (b) NEU-CLS[1] | (c) MTD [9] | (d) AITEX [10] | (e) MVTec [11] |

FIGURE 1. Examples of defects of various materials in datasets.

B. PREVIOUS SURVEYS

As an important research area, there have been already many reviews researching it. Table. 1 enlists some existing surveys in the industrial anomaly detection field, which have different focus from our research. Specifically, the review [73] starts with formulations of the most classical algorithms of different schools. The paper [74] surveys traditional methods, and also introduces deep learning-based methods. For each type of method, the characteristics of each method are listed in a general way. The article [75] represents an introduction to traditional methods, deep learning-based methods, and an introduction to hardware and software devices is involved in the meantime. For the introduction to deep learning-based methods, it mainly focuses on the supervised domain and introduces milestone algorithms by timeline. The survey [76] groups the relevant approaches given their underlying principles and discusses their assumptions, advantages, and disadvantages. The study [77] focuses on specific solutions for visual processing methods and, in particular, visual inspection approaches for metallic, ceramic, and textile surfaces in industrial applications. The methods in literature [78] are divided into categories based on the types of detection materials used. Some studies focus only on anomaly detection on a particular material. A thorough survey [1] is provided of both two-dimensional and three-dimensional surface defect detection systems for various common metal planar material products such as steel, aluminum, copper plates, and strips. The review [79] presents a detailed overview of histogram-based approaches, color-based approaches, image segmentation-based approaches, frequency domain operations, texture-based defect detection, sparse feature-based operations, and image morphology operations for fabric defect detection. The article [80] investigates supervised and semi-supervised deep learning algorithms. As for the unsupervised aspect, more attention is paid to the different network architectures used

by different types of algorithms. The article [81] focuses on the intersection of different research fields, providing extended cross-cutting ideas, exhaustively introducing the algorithms and frameworks of some typical methods, de-emphasizing methodological schools and blurring domain boundaries. It intends to bring these fields closer together. The article [82] focuses only on GAN-based algorithms.

In practice, it is more biased towards the needs for the unsupervised domain in the current industrial context. To our best knowledge, no review has been done for the recently emerged unsupervised methods. The article will provide a comprehensive and in-depth summary of the state-of-the-art algorithms for visual industrial anomaly detection, which will be divided into a systematic categorization listed as III-A Reconstruction-based, III-B Normalizing Flow (NF)-based, III-C Representation-based, III-D Data augmentation-based, and III-E Algorithm enhancements. This comprehensive summary is expected to contribute to the implementation and practice of the industrial field.

III. METHODOLOGY

A large part of the traditional visual anomaly detection algorithms belong to the category of supervised learning [83], [84], which requires collecting enough samples of different defect categories and accurate labeling, such as the category of the image, the location of the defects in the image and the category information of each pixel. However, in many application scenarios, it is difficult to collect a sufficient number of samples [85]. For example, in the surface defect detection task, most of the images collected actually belong to normal defect-free samples, while only a small amount belong to defect samples. With diverse types of defects to be detected, the number of defect samples available for training is very limited.

TABLE 1. Summary of previous reviews.

Tile	Year	Venue	Description
Deep learning for anomaly detection [79]	2020	KDD	This review combs **formulations** of the most classical algorithms of different schools.
A review on recent advances in vision-based defect recognition towards industrial intelligence [80]	2021	Journal of Manufacturing Systems	The reviews present a **concise overview** of traditional and deep learning-based visual anomaly detection techniques.
Recent advances in surface defect inspection of industrial products using deep learning techniques [81]	2021	Advanced Manufacturing Technology	
Visual Anomaly Detection for Images: A Systematic Survey [82]	2022	Procedia Computer Science	
Visual-based defect detection and classification approaches for industrial applications—a survey [83]	2020	Sensors	These studies focus only on anomaly detection on **particular materials**.
A review on industrial surface defect detection based on deep learning technology [84]	2020	MLMI	
Fabric defect detection using computer vision techniques: A comprehensive review [85]	2020	Mathematical Problems in Engineering	
Research progress of automated visual surface defect detection for industrial metal planar materials [1]	2020	Sensors	
Image-based surface defect detection using deep learning: A review [86]	2021	Computing and Information Science in Engineering	For the unsupervised aspect, more attention is paid to the different **network architectures** used by different types of algorithms.
A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges [87]	2021	arXiv	The article focuses on the **intersection** of different research fields, de-emphasizing methodological schools and blurring domain boundaries.
GAN-based Anomaly Detection: A Review [88]	2022	Neurocomputing	This paper focus on **GAN-based** anomaly detection and discusses its theoretical basis and applications.

Unsupervised visual anomaly detection algorithms can build detection models without any annotated samples, which makes it very suitable for the scenarios described above. In anomaly detection tasks, the difficulty in collecting normal images is much lower than that of anomalous images, which can significantly reduce the time and labor cost of detection algorithms in practical applications. Moreover, unsupervised visual anomaly detection models detect anomalous samples by analyzing the differences between normal samples and abnormal samples, allowing the algorithm to detect a wide range of abnormal samples, even brand new sorts of flaws. Comparison of framework diagrams of supervised and unsupervised algorithms is shown in Fig. 2.

There are some highlighted approaches worth mentioning among the algorithms with great performance, which will be described in detail in this section. We categorize the existing research into five types: reconstruction-based methods, normalizing flow-based methods, representation-based methods, data augmentation-based methods, and algorithm enhancement. An overview and summary of these categories is listed in Table 2.

FIGURE 2. Comparison of framework diagrams of supervised and unsupervised algorithms.

A. RECONSTRUCTION-BASED METHODS

To learn the distribution pattern of normal images, the core idea is to conduct encoding and decoding on the input normal images and train the neural network with the aim of reconstructio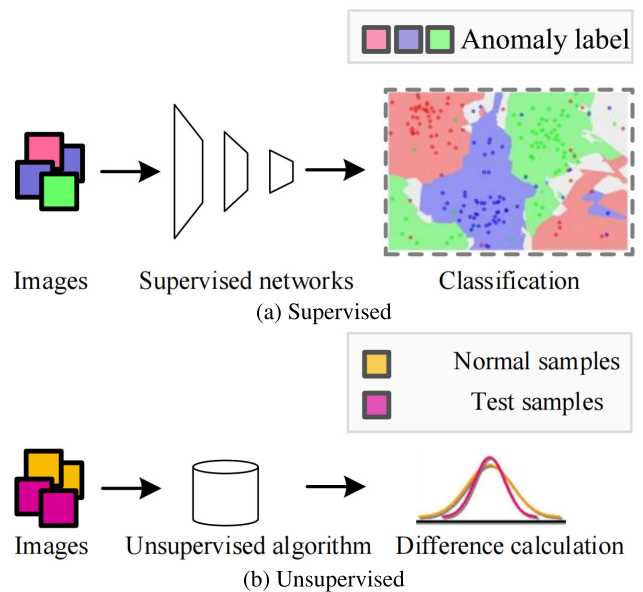n. With the help of the trained networks, the differences between the images before and after reconstruction are analyzed to detect anomalies in the detection stage. With anomaly score usually represented by reconstruction error, the anomalous images are easy to be found because

TABLE 2. An overview and summary of different categories of anomaly detection algorithms.

| Category	Description	Paper		Advantages	Disadvantages
Reconstruction-based Methods	The core idea is to conduct encoding and decoding on the input normal images and train the neural network with the aim of reconstruction. Based on the assumption that by training only on normal images, the model will not be able to reconstruct abnormal images correctly, and the anomaly scores will be higher.	Ven.,CAVGA Liu,UTAD Yang,DFR Mo.,STPM Yamada,RSTPM Deng,RDOE Rudolph,AST Massoli,MOCCA Liang,OCR-GAN Mishra,VT-ADL Li,SOMAD [102]	[92]; [93]; [94]; [95]; [96]; [97]; [98]; [99]; [100]; [101];	The algorithm principle is straightforward and comprehensible, while the network architecture is uncomplicated.	Reconstructors like autoencoders and GANs are highly generalizable and robust, and the anomalous part of the image can be well reconstructed, leading to hypothesis failure.
Normalizing Flow (NF)-based Methods	NF is able to learn transformations between data distributions and well-defined probability density functions, which can serve as a suitable estimator of probability densities for the purpose of detecting anomalies.	Rudolph,Differnet Gudovskiy,CFlow Rudolph,FCCSF Yu,FastFlow [106]	[103]; [104]; [105];	NF mapping is bijective and can be evaluated in both directions and inference efficiency is high.	NF-based methods require expensive training computational resources.
Representation-based Methods	Deep neural networks are used to extract meaningful vectors describing the image, and the anomaly score is usually represented by distance calculation.	Cohen,SPADE Defard,PaDIM Kars.,PatchCore Wang,GP Zheng,FYD Rip.,Gaussian-AD Yi,Patch SVDD DisAug CLR Ki.,Semi-orth Lee,CFA [116]	[107]; [108]; [109]; [110]; [111]; [112]; [113]; [114]; [115];	Representation-based methods do not call for a dedicated training stage, which introduces no parameters other than the backbone.	Because the backbone is usually biased towards ImageNet, it does not have good generalizability for different domains.
Data augmentation-based Methods	Augmentation algorithms are designed to resemble anomalies.	Zav.,DRAEM Schluter,NSA Li,CutPaste Nicolae,SSPCAB Song,AnoSeg [121]	[117]; [118]; [119]; [120];	The approach is uncomplicated and facile to comprehend and execute.	The data augmentation method is unable to fully replicate actual anomalies, resulting in certain generalization issues.
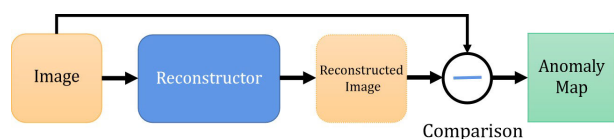
FIGURE 3. The basic flow of reconstruction-based method. The image is input into a reconstructor such as AE and GAN, to output the reconstructed image and compare it with the original input image. The difference between the two is used to get an anomaly map so as to realize anomaly detection.

they cannot be reconstructed well. Classical methods based on reconstruction include autoencoders (AE [116], [117]), variational auto-encoders (VAE [118]) and generative adversarial networks (GAN [119]), which can generate samples from the manifold of the training data. During the training phase, only normal data without anomalies are conventionally modeled. In testing phase, anomaly scores are calculated with the difference between the input image and the reconstructed image. Based on the assumption that by training only on normal images, the model will not be able to reconstruct abnormal images correctly, and the anomaly scores will be higher. The basic flow of reconstruction-based method is shown in Fig. 3. Typical auto-encoder and GAN based anomaly detection approaches mentioned above have some limitations, including:

Uncertain threshold. Autoencoders and GAN based approaches use a thresholded pixel-wise difference between the input and reconstructed image to localize anomalies. However, the use of anomalous training images, which may not be available in real-world situations, is required for these approaches to determine class-specific thresholds.

High computational cost. Autoencoder and GAN based anomaly detection approaches often require large amounts of computational resources to train and evaluate. This can be a bottleneck for real-time applications that require fast anomaly detection.

Difficult to interpret. The representations learned by autoencoders and GANs may be difficult to interpret, making it challenging to understand why a particular instance was classified as an anomaly. This can make it difficult to diagnose and fix problems in the system.

Sensitivity to hyperparameters. The performance of autoencoder and GAN based anomaly detection approaches can be sensitive to the choice of hyperparameters, such as the number of layers, the learning rate, and the batch size. It can be challenging to select the optimal hyperparameters for a given dataset, and the performance can degrade significantly if the hyperparameters are not tuned properly.

To improve the reconstruction ability, in method CAVGA [86], VAE, GAN and other means are combined and an attention mechanism is introduced for the first time
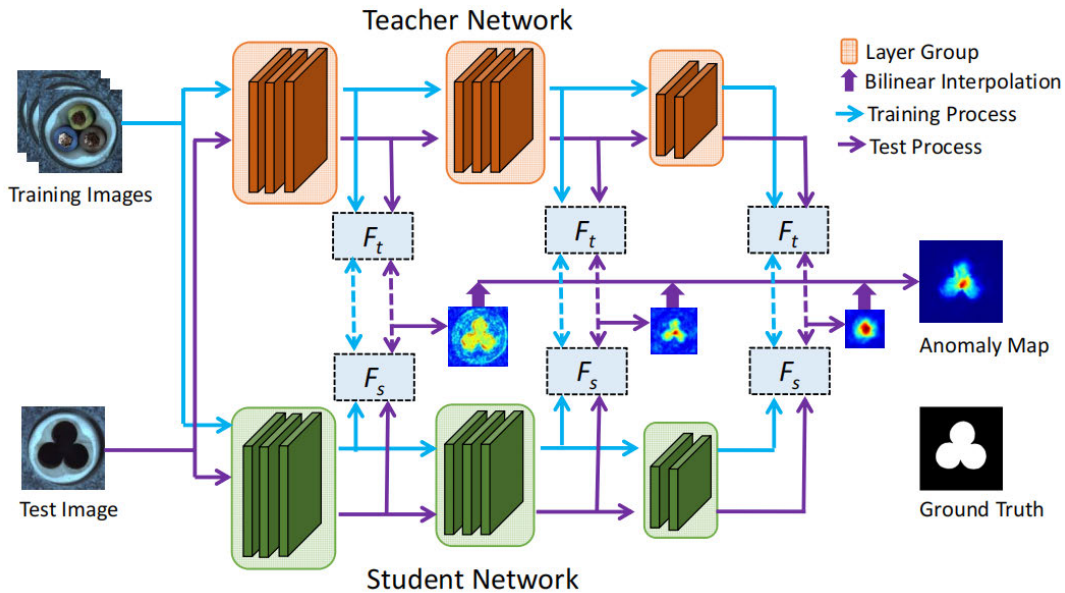
FIGURE 4. Schematic overview of STPM [89]. A student network's feature pyramid is trained to match the counterpart of a pre-trained teacher network. If the features from the two models disagree significantly, a test image (or pixel) gets a high anomaly score. STPM approach can detect anomalies of various sizes with a single forward pass owing to the feature pyramid matching scheme.

into the anomaly detection field. The framework encourages the attention map to cover the entire normal region, while suppressing attention maps corresponding to the anomaly classes in the training images. Two modes of unsupervised and weakly supervised are provided. 1. Unsupervised mode: GAN is used as the overall architecture, VAE is used as the codec and attention map is generated by Grad-CAM. Loss function consists of three parts: VAE, adversarial loss, and attention part. 2. Weakly supervised mode: Compared with mode 1, classifiers are added to distinguish normal and abnormal samples. Loss function consists of four parts: VAE, adversarial loss, complementary guided attention loss, and classification loss. However, one potential drawback is that it relies on the assumption that anomalies in images are always visually distinct from the background or normal regions. This may not always be the case, as certain types of anomalies may be visually similar to the background, making them harder to detect using this method.

Classical methods like GAN and Autoencoder compare the input and its difference from the output to pinpoint the anomaly. However, coarse reconstructions produce excessive image differences, which prevents the detection of anomalies. To address this problem, the approach UTAD [87] proposes an unsupervised visual anomaly detection method for natural images by combining mutual information, GAN, and autoencoder. A two-stage framework (i.e., IE-Net, Expert Net) is utilized to generate high-fidelity and anomaly-free input reconstructions for anomaly detection tasks.

Aiming at the anomalies in small and confined regions of images, DFR method [88] suggests an effective unsupervised anomaly segmentation approach, which utilizes the transformed hierarchical CNN features to build dense discriminative multiscale feature representations for every local region of the images via a specially designed regional feature generator. DFR also proposes to detect possible anomalous regions in images by deep feature reconstruction, i.e. reconstructing the multiscale regional features via a deep yet efficient convolutional autoencoder(CAE). The regional feature generator takes the multi-scale feature maps as input and turns them into a relatively large single feature map, which is then reconstructed by a deep CAE. By calculating the reconstruction error and the anomaly score map, anomalies will be segmented if any score on the anomaly map is greater than the estimated value or a user-defined threshold.

Some attempts utilize pre-trained model of image classification task. Nevertheless, the problem of the incompleteness of transferred knowledge and the complexity of handling scaling has not yet been resolved. Thus, STPM [89] introduces a novel feature pyramid matching technique and incorporates it into the student-teacher anomaly detection framework. Fig. 4 shows the overview of STPM. The algorithm employs multiple layers of features extracted from a powerful network pre-trained for image classification tasks as the teacher to guide a student network with the same structure to learn the distribution of anomaly-free images. The student network learns the distribution of images by matching the features of the anomaly-free images with the pre-trained network, and this step of transmission seeks to retain as much critical information as possible. In the training phase, the teacher network is a mature network trained on ImageNet, and the image input network generates multi-layer feature maps. The student network is trained with a fraction of the

training set, approximating the multi-layer feature trained by the teacher network as much as possible. In the testing phase, samples are put into both teacher and student networks, the differential loss between which is computed. A high anomaly score will be assigned if the features of a test image (or pixel) deviate significantly between the two models. If any pixel in the image is anomalous, the image is judged as anomalous.

The RSTPM [90] approach is a generalization of the Student-Teacher framework method STPM, which was developed previously. The new approach differs from prior STPM in three ways: student network for reconstruction, an attention mechanism from the teacher network to the student network, and a different teacher network structure from the original STPM.

T-S model typically uses similar or identical architectures. To improve the T-S model's representation diversity on unknown, out-of-distribution samples, a novel T-S model [91] with a teacher encoder and student decoder is suggested, along with a straightforward yet powerful reverse distillation paradigm. Instead of receiving raw images directly, the student network takes the teacher model's one-class embedding as input and targets to restore the teacher's multiscale representations. It is the first approach to adopt an encoder and a decoder to construct the T-S model. This strategy differs from existing ones due to the heterogeneity of the teacher and student networks and reversed data flow in knowledge distillation.

Previous methods suffer from the similarity of student and teacher architecture, such that the distance is undesirably small for anomalies. To tackle this problem, AST [92] proposes asymmetric student-teacher networks, which train a normalizing flow for density estimation as a teacher and a conventional feed-forward network as a student to trigger large distances for anomalies.

Explicitly leveraging the networks' multi-layer composition, MOCCA [93] exploits the output of a deep model at different depths to detect anomalous input in the one-class setting. With MOCCA, the training technique is split into two stages in which the autoencoder is trained on the reconstruction task only, and then only the encoder is utilized to detect anomalies by exploiting a one-class-like objective applied to different layers of the network.

OCR-GAN [94] reconsiders the distinction between normal and abnormal images from the frequency domain perspective and proposes a novel framework for anomaly detection based on omni-frequency reconstruction. Specifically, FD module is proposed to decouple the input image into various frequencies and model the reconstruction process as a combination of parallel omni-frequency image restorations.

VT-ADL [95] combines the classic reconstruction-based methods with the benefits of a patch-based approach. Visual transformer networks contribute to preserving the spatial information of the embedded patches, which is later coped with a Gaussian mixture density network to localize the anomalous areas.
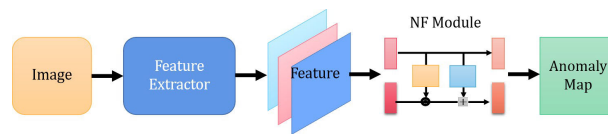

FIGURE 5. The basic flow of the Normalizing Flow (NF)-based method. The features are first extracted by a feature extractor and then fed into NF module to estimate the probability density. The probability value is employed as the anomaly score in testing phase.

Self-organizing map for anomaly detection (SOMAD) [96] makes use of pre-trained CNN to extract the features of patches and leverages the SOM to maintain the neighborhood relationship of embedding vectors in topology space. It greatly reduces the search space by mapping the normal feature space into 2-dimensional space through SOM.

B. NORMALIZING FLOW (NF)-BASED METHODS

Normalizing Flows (NF) [120] are neural networks that are able to learn transformations between data distributions and well-defined probability density functions. Their special property is that their mapping is bijective and they can be evaluated in both directions. The property of normalizing flows to serve as a suitable estimator of probability densities for the purpose of detecting anomalies has not drawn much attention yet. Here we summarize some recently emerged NF-based visual anomaly detection algorithms to provide ideas for future study.

There are methods [97], [98] adopting normalizing flow to estimate distribution through a trainable process that maximizes the log-likelihood of normal image features. Normal image features are embedded into standard normal distribution and the probability is used to identify and locate anomalies. The basic flow of the Normalizing Flow (NF)-based method is shown in Fig. 5.

The method [99] detects and locates defects based on density estimates of feature maps extracted from input images of different sizes. Cross-connections between scales are made by jointly processing multiscale feature maps using a fully convolutional normalizing flow.

However, in order to estimate the distribution, the original one-dimensional normalizing flow model must flatten the two-dimensional input feature into a one-dimensional vector, which destroys the inherent spatial positional relationship implied by the two-dimensional image and constrains the NF model. FastFlow [100] expands the original normalizing flow model to two-dimensional space to address the concerns mentioned above. As shown in Fig. 6, the algorithm is summarized: the visual features are first extracted by a feature extractor and then fed into FastFlow to estimate the probability density. In the training phase, FastFlow is trained with normal images to transform the original distribution into a standard normal distribution in a 2D manner. In inference phase, the probability value of each position on the 2D feature is employed as the anomaly score.
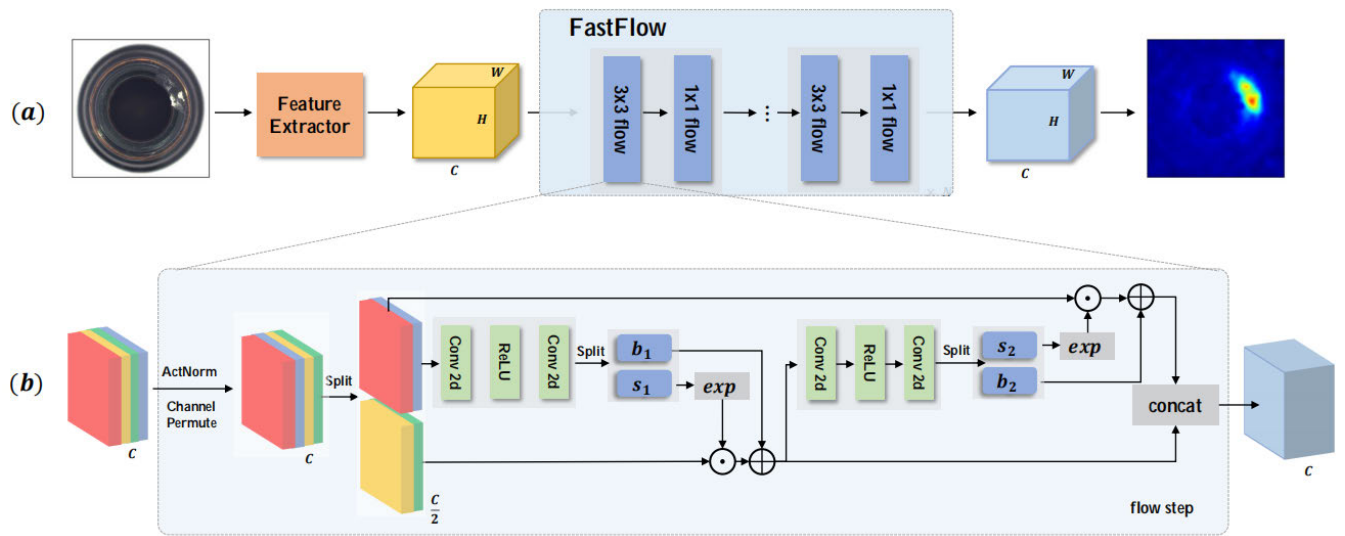
FIGURE 6. (a) shows the whole framework of FastFlow [100] algorithm. (b) is one flow step of FastFlow.
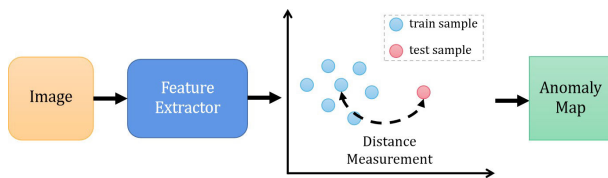

FIGURE 7. The basic flow of representation-based method. Pre-trained deep neural networks are used as feature extractor to extract meaningful vectors describing the input image, and the anomaly map is usually represented by the distance between the test embedded vectors and the reference vector representing normality from the training dataset.

C. REPRESENTATION-BASED METHODS

For representation-based methodology, deep neural networks are used to extract meaningful vectors describing the entire image, and the anomaly score is usually represented by the distance between the embedded vectors of the test images and the reference vector representing normality from the training dataset. The basic flow of representation-based method is shown in Fig. 7. The core idea is to train a deep neural network as a feature extractor to make the distribution of feature vectors extracted from normal images as compact as possible, i.e., the intra-class distance of the samples is reduced as much as possible. Contrary to reconstruction-based algorithms, representation-based methods do not call for a dedicated training stage, which introduces no parameters other than the backbone. The concept of distance metric learning techniques is comparable to clustering.

In the testing phase, most methods calculate the distance between the features of the sample to be tested and the normal features as a metric to perform anomaly detection. Typical algorithms mainly include SPADE [101], PaDiM [102], PatchCore [103], GP [104], etc. To record anomaly score and generate a score map, all these approaches employ different distance measurements (loss functions). PatchCore

uses a maximally representative memory bank of nominal patch-features to integrate embeddings from ImageNet models with an outlier detection model. The framework of Patch-Core [103] is shown in Fig. 8

PaDiM makes use of a pre-trained convolutional neural network (CNN) for patch embedding, and multivariate Gaussian distributions to get a probabilistic representation of the normal class. It also exploits the correlations between different semantic levels of the CNN to better locate the anomalies.

Based on alignment between an anomalous image and a constant number of similar normal images, SPADE [101] uses KNN and multiscale feature pyramid for defect detection and localization of anomalies. The following steps make up SPADE algorithm: i) image feature extraction ii) K-nearest-neighbor normal image retrieval iii) pixel alignment with deep feature pyramid correspondences.

With a coarse-to-fine alignment technique, FYD method [105] seeks to learn dense and compact distribution of normal images. In both picture and feature levels, the coarse alignment stage normalizes the pixel-level position of objects. After that, the fine alignment stage maximizes the similarity of features across all corresponding locations in a batch.

In terms of the scale of image processing, the methods can be divided into image level, patch level and pixel level. Gaussian-AD [106] extracts discriminative feature vectors from normal images. Algorithms like Patch SVDD [107], PatchCore [103] and PaDiM [102] extract discriminative feature vectors from normal image patches. SPADE [101] extracts discriminative features which are used for pixel-level image alignment. From different process levels, these methods extract features of normal images and model the distribution with statistical methods. Based on the assumption that abnormal samples have different distributions, more promising results for anomaly detection are yielded.
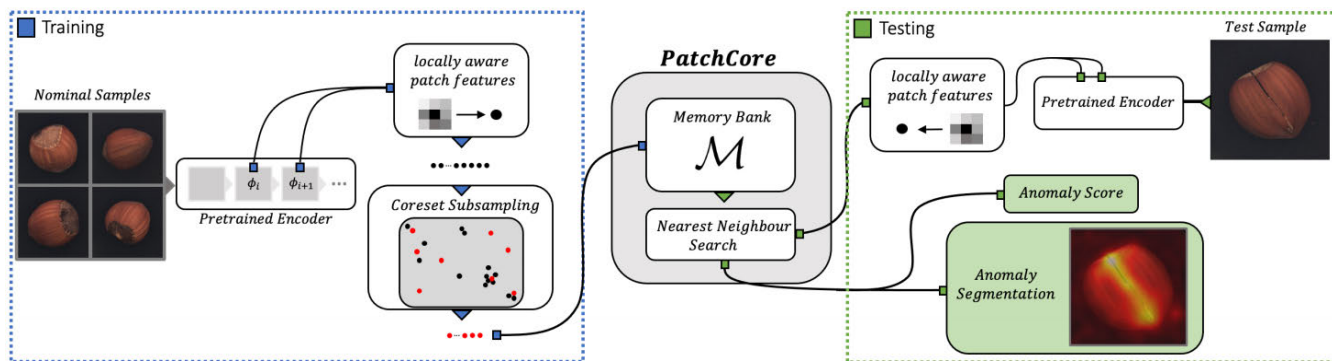
FIGURE 8. The overall framework of PatchCore. During training, normal samples are decomposed into a memory bank of neighborhood-aware patch-level features. To reduce redundancy and inference time, this memory bank is downsampled via greedy coreset subsampling algorithm. During test time, images are classified as anomalies if at least one patch is anomalous, and pixel-level anomaly segmentation is generated by scoring each patch feature.

Based on distribution-augmented contrastive learning, DisAug CLR algorithm [108] first learns self-supervised representations from one-class data, and then builds one-class classifiers on learned representations.

The method Semi-orthogonal [109] is a generalization of the prior work's random feature selection method PaDIM [102]. It extends the random feature selection to semi-orthogonal embedding as a low-rank approximation of precision matrix for the Mahalanobis distance.

Albeit simple and efficient, most of these methods require manual specification of feature centers in advance, and additional tasks need to be designed in the training stage to avoid model degradation. The approach of setting only one global feature centroid imposes some constraints on the image context. In the changeable scenes of medical images or natural images, it may be challenging to map all images to the same target point under the condition of guaranteeing generalization ability.

Previous studies focused on approximating the distribution or extracting features with pre-trained CNNs of normal data, which may make the normality of abnormal features overestimated. CFA [110] performs transfer learning on the target dataset as a solution to alleviate this problem. CFA first acquires multiscale feature maps with biased CNN to generate a patch memory bank. Through transfer learning and the feature adaptation of patch descriptor associated with the memory bank, CFA achieved successful target-oriented anomaly detection.

D. DATA AUGMENTATION-BASED METHODS

In the unsupervised setting, the training data are all anomaly-free data. Hence, there are some algorithms [108], [111], [112], [113], [115] that adopt the method of creating anomalies. To overcome the limitation of insufficient data, augmentation algorithms [121], [122] have been widely used in deep learning scheme. The basic flow of the data augmentation-based method is shown in Fig. 9.

In DRAEM [111] method, noise generation method is adopted to create anomalies and superimpose them on normal
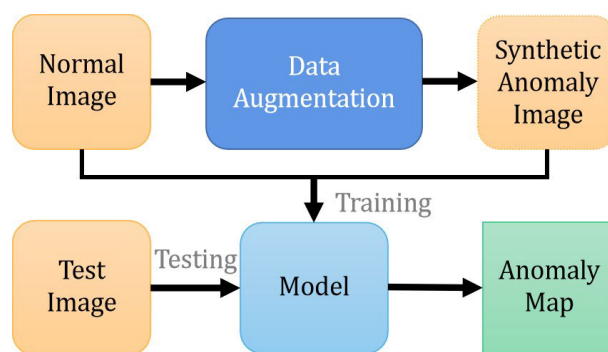

FIGURE 9. The basic flow of the data augmentation-based method. The normal image is augmented to obtain synthetic abnormal image. Both normal samples and synthetic abnormal samples are input into the model for training. The difference generated after the test sample input into the model is regarded as anomaly map.

images. The proposed method learns a joint representation of a normal image and its synthetic anomalous image, while simultaneously learns a decision boundary between normal and anomalous examples.

The method NSA [112] is a naturally synthetic anomaly approach that proposes a way to create anomalies by selecting patches of different sizes at different locations and blending them into anomaly-free images. Specifically, it is a self-supervised task to create diverse and realistic synthetic anomalies with Poisson image editing to seamlessly blend multiscale patches of various sizes in different images. This produces a wide range of synthetic anomalies, which are more similar to natural sub-image irregularities.

CutPaste [113] is also an synthetic anomaly method designed to produce augmentations to synthesize anomalous samples by operating on normal image patches, including cropping, rotating, transforming and overlaying. The distance between the normal samples and the generated anomalous samples is then measured. An overview of CutPaste method for anomaly detection and localization is shown in Fig. 10.

The method [114] proposes a self-supervised predictive convolutional attentive block (SSPCAB), which can be easily
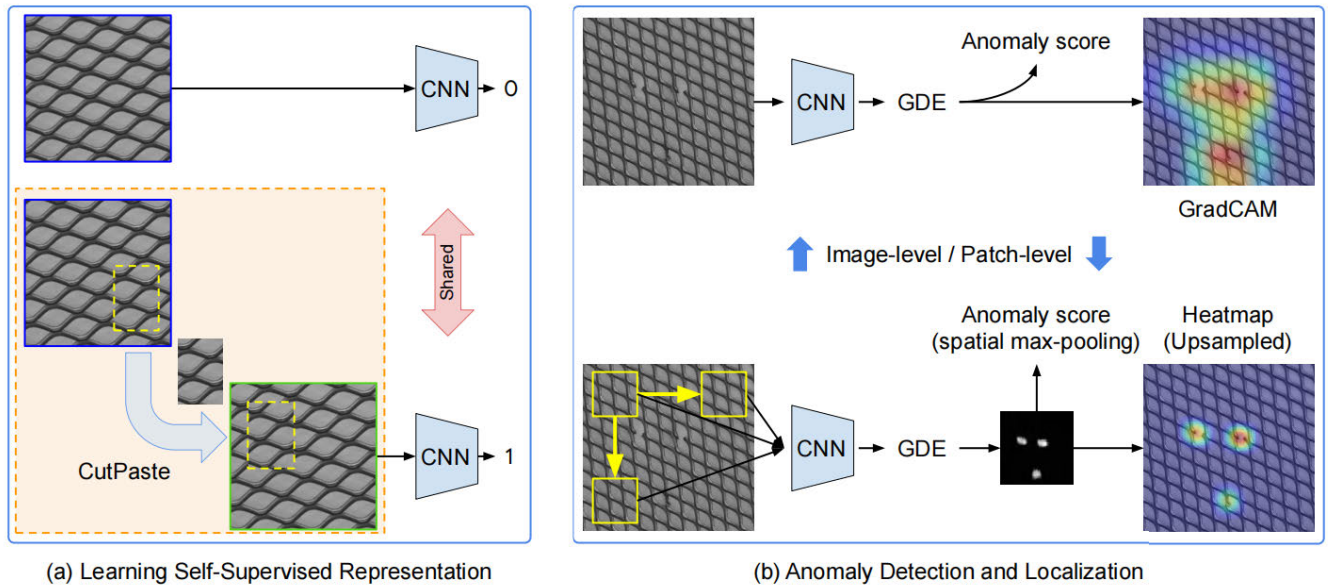
(a) Learning Self-Supervised Representation

(b) Anomaly Detection and Localization

FIGURE 10. Overview of CutPaste [113]. (a) The CutPaste method creates an abnormal image by cropping a random area of the defect-free image and superimposing it on the normal image. Then the CNN classifier is trained jointly with normal samples and synthetic anomalous samples. (b, top) Image-level inference process. Gaussian density estimator(GDE) is adopted to compute anomaly scores via one-class classifiers. (b, bottom) Patch-level inference process. Features are extracted from local patches to produce anomaly score map, which is then max-pooled for detection or upsampled for localization.

incorporated into various state-of-the-art anomaly detection methods, such as DRAEM [111] and CutPaste [113]. It aims at reconstructing masked information with contextual information, so as to realize performance improvements.

AnoSeg [115] is a segmentation model which combines three techniques: self-supervised learning with hard augmentation, adversarial learning, and coordinate channel connectivity. It is directly trained for anomaly segmentation tasks with synthetic anomaly data generated by hard augmentation. In addition, anomaly regions sensitive to positional relationships are more easily to be detected by means of coordinate vectors representing the pixel position information.

E. ALGORITHM ENHANCEMENTS

Some algorithms provide some enhancements [107], [123], [124], [125], such as improved loss functions or interpretability.

IGD [123] employs reverse-interpolated training samples to train a class of Gaussian anomaly classifiers that describe representative normal samples for effective normality. Current state-of-the-art models learn a compact normality description by hyper-sphere minimization, but they are prone to overfitting. To solve this problem, interpolated Gaussian descriptor (IGD) approach is introduced. Methods that can locate anomalies generally are suitable for a specific anomaly size and structure, which may result in missing anomalies outside of that size and structure range. To avoid this problem, IGD is designed to detect multiscale structural and non-structural anomalies to improve the accuracy of anomaly localization.

Classical unsupervised anomaly detection algorithms such as support vector data description (SVDD [126]) and Deep-SVDD (DSVDD [127]) can hardly explain why an image is anomalous. Therefore, FCDD [124] explores converting the final comparison vector of the previous DSVDD model into a two-dimensional matrix (explanation heatmap) to enhance the interpretability of the algorithm. For most traditional fully connected convolutional networks, images are mapped to the feature map of $1 * U * V$. It is mentioned in this paper that an important attribute of the convolution layer is that a pixel of the feature map only has a fixed receptive field corresponding to the input. A heatmap upsampling algorithm is proposed in this paper, so that the abnormal score of the feature map can be mapped back to the position of the original image, i.e., spatial information is reserved.

A new loss function is proposed which can overcome failure modes of both center-loss and contrastive-loss methods [125]. Furthermore, it is combined with a confidence-invariant center loss, which replaces the Euclidean distance used in previous work, i.e., a distance that is sensitive to prediction confidence. The improvements yield a new anomaly detection approach, based on mean-shifted contrastive loss, which is both more accurate and less sensitive to catastrophic model collapse than previous methods.

In the field of anomaly detection, attention mechanisms [86], [128] are often used for algorithm improvement. Another kind of methodology utilizes multiscale features to enrich semantic information capture [88], [89], [93], [97], [98], [99], [101], [102], [106].

There are other methods try brand new ways to solve anomaly detection task. For the first time, RFS Energy

TABLE 3. Comparison between supervised and unsupervised algorithms.

	Need label	Support for undefined defects	Accuracy	Generalizability
Supervised	✓	✗	higher	✗
III-A Reconstruction-based	✗	✓	high	✓
III-B NF-based	✗	✓	high	✓
III-C Representation-based	✗	✓	high	✗
III-D Augmentation-based	✗	✗	low	✗

TABLE 4. Complexity comparison in terms of inference speed (FPS), additional inference time (millisecond) and number of additional parameters (M) for various backbones. A.d. Time means the additional inference time and A.d. Parmas is the number of additional parameters compared with backbone network.

Model	FPS	A.d. Time (ms)	A.d. Params (M)
Patchcore [109] (Representation-based)	5.88	159	0
CFlow [104] (NF-based)	14.9	56	81.6
DFR [94] (Reconstruction-based)	100	43.8	124

algorithm [129] solves the challenge of unsupervised anomaly detection using keypoint detection and an energy model.

IV. COMPARISON AND ANALYSIS

Both supervised and unsupervised algorithms are used in the field of anomaly detection, and the advantages and disadvantages of each are summarized in Table 3. Although the supervised approach possesses high accuracy, there are limitations in the acquisition of labeled data, which requires a large amount of work, and sometimes it is impossible to acquire enough labeled defect samples. The process of training the network also has many parameters to optimize, which leads to inefficiency. Classification is not possible for defects that do not appear in the training set. The classes of methods introduced above in this paper are all unsupervised algorithms that do not require category labels, which can save a lot of cost and effort in practical applications. NF-based methods require expensive training computational resources, while undefined defect detection is supported and inference efficiency is high. Reconstruction-based methods require expensive training for the related task and deep generative models are not robust enough, and their performances for anomaly detection are not stable whereas the model has good generalization ability. Representation-based methods do not need to introduce parameters other than backbone, which is beneficial for efficiency. However, because backbone is usually biased towards ImageNet, it does not have good generalizability for some images, such as medical images. Data augmentation-based methods are designed to resemble the anomalies, which are data-dependent and non-automatic.

As far as complexity is concerned, we take time complexity and memory complexity into account. **Time complexity.** For representation-based algorithms, the training time complexity scales linearly with the dataset size. However,

contrary to the methods that require training deep neural networks like reconstruction-based methods, representation-based algorithms use a pre-trained CNN, and, thus, no deep learning training is required which is often a complex procedure. Hence, it is very fast and easy to train on small datasets like MVTec AD. Conversely, take SPADE as an example, it computes and stores in the memory before testing all the embedding vectors of the normal training images. Those vectors are the inputs of a KNN algorithm which makes SPADE's inference speed very slow. While for reconstruction-based methods, after training stage, their inference phase can be quite fast. NF-based methods avoid the time-consuming k-nearest-neighbor-search process, while it still needs to perform a more complex inference phase than reconstruction-based methods. **Memory complexity.** Representation-based algorithms like SPADE and Patchcore perform KNN clustering between each test feature of each image patch and the gallery features of normal image patches, and they do not need to introduce parameters other than backbone. But they require large memory allocation for gallery features.

We make an efficiency analysis of some representative methods from aspects of inference speed, additional inference time and additional model parameters, "additional" refers to not considering the backbone itself. The hardware configuration of the machine used for testing is Intel(R) Xeon(R) CPU E5-2680 V4@2.4GHZ and NVIDIA GeForce GTX 1080Ti. The analysis results are shown in Table 4.

V. DATASETS

Datasets are the base for research work. A good dataset is more conducive to the discovery and summary of problems, so as to facilitate the solution. There are now some quality inspection/anomaly detection datasets in the industry field.

A. BTAD

BeanTech Anomaly Detection dataset [95] (BTAD[1]) contains a total of 2830 real-world images of 3 industrial products showcasing body and surface defects. The training set consists of only normal images, while the testing set has a mixture of both normal and abnormal images. Product 0, 1, and 2 of this dataset contain 400, 1000, and 399 training images respectively. This dataset is often used for unsupervised defect/anomaly detection. The AUROC (area under the receiver operator curve) metrics of the SOTA methods on this dataset are summarized in Table 5, where the bold parts are the best-performing results.

B. SOLAR PANEL DATASET: ELPV

The dataset ELPV[2] [133], [134], [135] contains 2624 8-bit grayscale image samples of 300 × 300 pixel functional and defective solar cells, with varying degrees of degradation extracted from 44 different solar modules. Defects in annotated images are internal or external types of defects known

---

[1] http://avires.dimi.uniud.it/papers/btad/btad.zip

[2] https://github.com/zae-bayern/elpv-dataset

TABLE 5. Anomaly localization results measured by pixel-wise AUROC on BTAD dataset.

Methods	Product0	Product1	Product2	Average
VT-ADL [101]	**0.990**	0.940	0.770	0.900
AE MSE [136]	0.490	0.920	0.950	0.787
AE MSE+SSIM [137]	0.530	0.960	0.890	0.793
FastFlow [106]	0.950	0.960	0.990	0.967
BGAD [138]	0.972	0.967	0.996	0.978
BGAD-FAS [138]	0.980	**0.977**	**0.998**	**0.985**
FYD [111]	0.961	0.953	0.997	0.970

TABLE 6. Image-level anomaly detection AUROC results on the AITEX and ELPV datasets.

Methods	AITEX	ELPV
KDAD [139]	0.576	0.744
DevNet [140]	0.598	0.514
FLOS [141]	0.538	0.457
SAOE [142]	0.675	0.635
MLEP [143]	0.564	0.578
DRA [144]	0.692	0.675
BGAD-FAS [138]	**0.826**	**0.903**

to reduce the power efficiency of solar modules. With every image annotated with a defect probability (a floating point value between 0 and 1), this dataset can be used to solve unsupervised tasks.

C. FABRIC DEFECT DATASET: AITEX
The collection AITEX[3] [136] contains photos of seven different fabric textures with a resolution of 4096 × 256 pixels. There are 140 defect-free images in the dataset, 20 images for each type of fabric. In addition, there are 105 images of 12 different types of fabric defects commonly found in the textile industry. It can be used to solve unsupervised tasks. The AUROC metrics of the SOTA methods on ELPV dataset and AITEX dataset are summarized in Table 6, where the bold parts are the best-performing results.

D. MTD-SURFACE DEFECT SALIENCY
In magnetic brick surface defect dataset[4] [143], a total of 1344 images are taken, the ROI (region of interest) of the tiles is cropped and classified into six subsets according to defect type, which are respectively porosity, crack, wear, fracture, non-uniformity (caused by the grinding process) and free (defect-free), each with a pixel-level label. To simulate the manufacturing process on an actual assembly line, images are captured under a variety of lighting conditions for each given brick. It can be used to solve unsupervised tasks. The experiment results of AUROC on this dataset are summarized in Table 7, where the bold parts are the best-performing results.

E. KolektorSDD
KolektorSDD [150] consists of 399 images of electrical commutators, where 52 defected images are annotated for

TABLE 7. Experiment results of AUROC on MTD dataset.

Methods	AUROC
Geom [145]	0.755
GANomaly [146]	0.766
DSEBM [147]	0.572
Mahalanobis [112]	0.980
1-NN [148]	0.800
DifferNet [103]	0.977
PaDiM [108]	0.987
CS-Flow [105]	**0.993**
PatchCore [109]	0.979
ADGAN [149]	0.464
OCSVM [150]	0.587

TABLE 8. AUROC performance on KolektorSDD dataset.

Methods	AUROC
skipGAN [152]	0.551
Puzzle AE [153]	0.554
DifferNet [103]	0.849
InTra [134]	0.701
CutPaste [119]	0.602
Draem [117]	0.859
OCR-GAN [100]	0.914
Uninformed student [154]	0.896
PaDiM [108]	0.945
Semi-orthogonal [115]	**0.960**

TABLE 9. AUROC performance on DAGM dataset.

Methods	Average
skipGAN [152]	0.558
Puzzle AE [153]	0.593
CutPaste [119]	0.665
DifferNet [103]	0.746
Draem [117]	0.980
OCR-GAN [100]	**0.993**
f-AnoGAN [155]	0.575
Uninformed student [154]	0.864
Staar [156]	0.830

microscopic fractions or cracks on the surface of the plastic embedding in electrical commutators. The dataset represents a real-world problem of surface-defect detection for an industrial semi-finished product where the number of defective items available for the training is limited. Table 8 shows AUROC performance on KolektorSDD dataset in terms of several SOTA algorithms.

F. DAGM
DAGM[5] [156] is a well-known benchmark dataset for surface defect detection. It contains images of various surfaces with artificially generated defects. Surfaces and defects are split into 10 classes of various difficulties, such as scratches or spots. It is a weakly supervised dataset, and there are 8,050 training and testing sets each, and the ratio of positive and negative samples for each type is approximately 1:7. The experiment results of AUROC on this dataset averaged over ten categories are summarized in Table 9, where the bold part is the best-performing result.

TABLE 10. Anomaly detection and localization performance on MVTec AD dataset with the format (image-level AUROC, pixel-level AUROC).

|	PatchSVDD [113]	SPADE [107]	CutPaste [119]	PatchCore [109]	AST [98]	CFA [116]	CFlow [104]	FastFlow [106]
carpet	(92.9,92.6)	(98.6,97.5)	(100.0,98.3)	(98.7,98.9)	(97.5,-)	(97.3,-)	(100.0,99.3)	(100.0,99.4)
grid	(94.6,96.2)	(99.0,93.7)	(96.2,97.5)	(98.2,98.7)	(99.1,-)	(99.2,-)	(97.6,99.0)	(99.7,98.3)
leather	(90.9,97.4)	(99.5,97.6)	(95.4,99.5)	(100.0,99.3)	(100.0,-)	(100.0,-)	(97.7,99.7)	(100.0,99.5)
tile	(97.8,91.4)	(89.8,87.4)	(100.0,90.5)	(98.7,95.6)	(100.0,-)	(99.4,-)	(98.7,98.0)	(100.0,96.3)
wood	(96.5,90.8)	(95.8,88.5)	(99.1,95.5)	(99.2,95.0)	(100.0,-)	(99.7,-)	(99.6,96.7)	(100.0,97.0)
bottle	(98.6,98.1)	(98.1,98.4)	(99.9,97.6)	(100.0,98.6)	(100.0,-)	(100.0,-)	(100.0,99.0)	(100.0,97.7)
cable	(90.3,96.8)	(93.2,97.2)	(100.0,90.0)	(99.5,98.4)	(98.5,-)	(99.8,-)	(100.0,97.6)	(100.0,98.4)
capsule	(76.7,95.8)	(98.6,99.0)	(98.6,97.4)	(98.1,98.8)	(99.7,-)	(97.3,-)	(99.3,99.0)	(100.0,99.1)
hazelnut	(92.0,97.5)	(98.9,99.1)	(93.3,97.3)	(100.0,98.7)	(100.0,-)	(100.0,-)	(96.8,98.9)	(100.0,99.1)
meta nut	(94.0,98.0)	(96.9,98.1)	(86.6,93.1)	(100.0,98.4)	(98.5,-)	(100.0,-)	(91.9,98.6)	(100.0,98.5)
pill	(86.1,95.1)	(96.5,96.5)	(99.8,95.7)	(96.6,97.1)	(99.1,-)	(97.9,-)	(99.9,99.0)	(99.4,99.2)
screw	(81.3,95.7)	(99.5,98.9)	(90.7,96.7)	(98.1,99.4)	(99.7,-)	(97.3,-)	(99.7,98.9)	(97.8,99.4)
toothbrush	(100.0,98.1)	(98.9,97.9)	(97.5,98.1)	(100.0,98.7)	(96.6,-)	(100.0,-)	(95.2,99.0)	(94.4,98.9)
transistor	(91.5,97.0)	(81.0,94.1)	(99.8,93.0)	(100.0,96.3)	(99.3,-)	(100.0,-)	(99.1,98.0)	(99.8,97.3)
zipper	(97.9,95.1)	(98.8,96.5)	(99.9,99.3)	(98.8,98.8)	(99.1,-)	(99.6,-)	(98.5,99.1)	(99.5,98.7)
Average	(92.1,95.7)	(96.2,96.5)	(97.1,96.0)	(99.1,98.1)	(99.2,95.0)	(99.2,98.2)	(98.3,98.6)	(99.4,98.5)
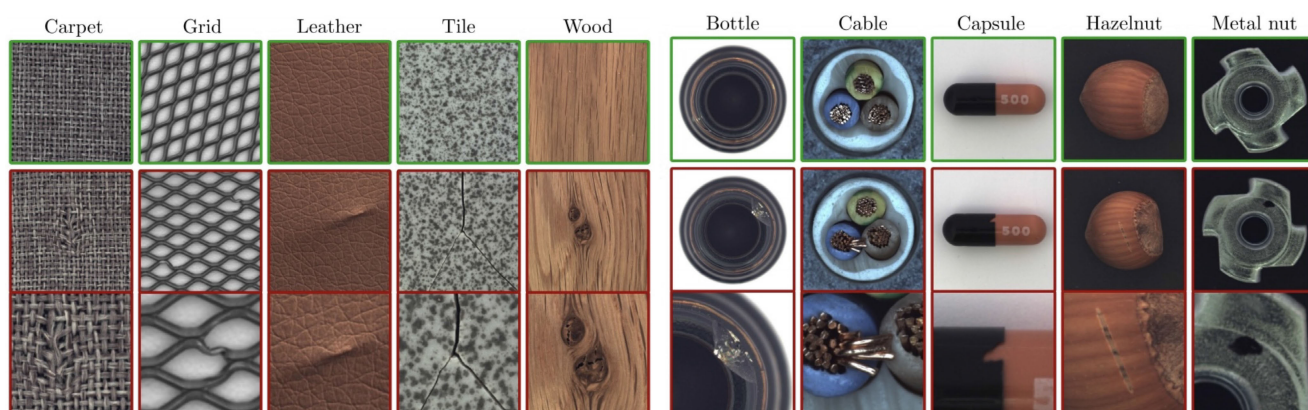
FIGURE 11. Example images of the MVTec AD dataset. For each category, the top row shows an anomaly-free image. The middle row shows an anomalous example. In the bottom row, a close-up view that highlights the anomalous region is provided.

TABLE 11. Challenges in anomaly detection. Different datasets illustrate the challenges in the industry anomaly detection field.

	BTAD	ELPV	AITEX	MTD-Surface	KolektorSDD	DAGM	MVTec AD
Small amount of anomalous data			✓		✓		
Small size of defects			✓			✓	
Object appearance variability	✓	✓					✓
Texture differences			✓	✓			

G. MVTec AD

MVTec AD dataset[6] [157] has a total of 15 categories, with 5 of them being distinct types of textures and the remaining 10 being different sorts of objects. In total, 3629 photos are utilized for training and verification, while 1725 images are used for testing in this dataset. The training set contains solely non-defective images, whereas the test set contains both non-defective and defective images of various types. This dataset is often used for unsupervised defect/anomaly detection. Example images of MVTec AD dataset are shown in Fig. 11. Under the metrics of image-level AUROC and pixel-level AUROC, the detailed comparison results of all categories are shown in Table 10.

[6]http://www.mvtec.com/company/research/datasets

H. SUMMARY

There are many challenges in visual industrial anomaly detection scenarios. Take the datasets we listed for example, as shown in Table 11, there are problems such as small amount of anomalous data, small size of defects, object appearance variability, texture differences, etc.

VI. CHALLENGES AND DISCUSSION
A. LACK OF COMPREHENSIVE OPEN DATASETS

Currently, the existing open datasets merely cover a limited number of scenarios, which is not comprehensive enough. The actual industrial scenarios are rich and diverse, resulting in a domain gap with the scenarios presented by the open datasets. Although the AUROC of existing methods on open
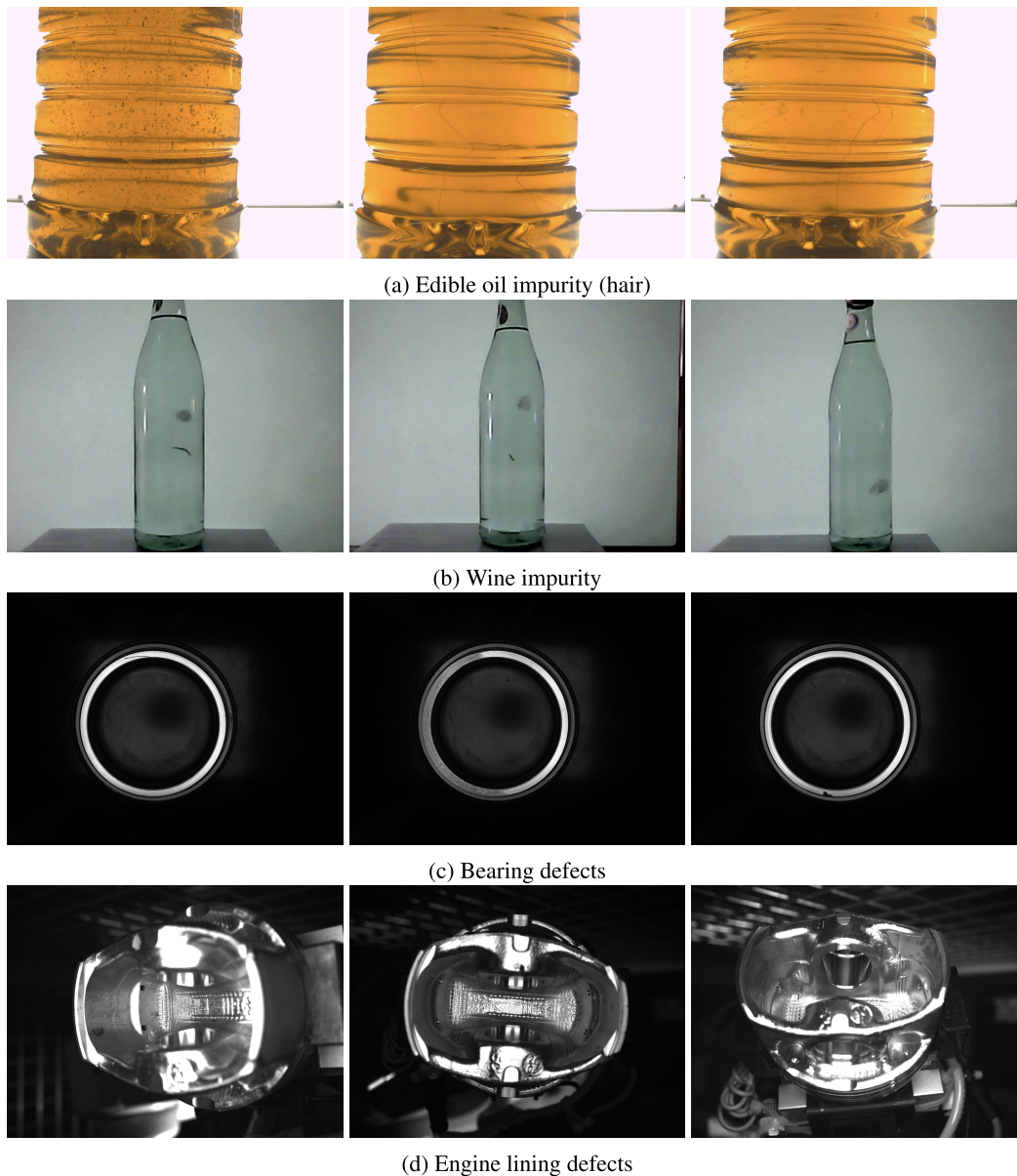
(a) Edible oil impurity (hair)

(b) Wine impurity

(c) Bearing defects

(d) Engine lining defects

FIGURE 12. Examples of defects of different industrial materials.

datasets is high, it is not sufficiently instructive. In industrial quality inspection scenarios, defects are complex and diverse, and the current study is only the tip of the iceberg. In the actual industrial scenario, there are also cases such as edible oil impurities, wine impurities, bearing defects, engine lining defects (3D internal), etc., as shown in Fig. 12. There are still some problems that are not well solved by current methods. Taking abnormal detection of edible oil in the actual industrial scene as an example, we verify the anomaly detection of image level with existing SOTA algorithm Patchcore and AST as shown in Table. 12, but the experimental results are not ideal and far from the results reported on MVTec benchmark. It further explains that the data in the actual industrial scenario is more complex, and the benchmark data is too simple and not rich enough. On the other hand, the lack of

TABLE 12. Experimental results of existing SOTA methods on edible oil data.

	Patchcore-single	Patchcore-3model	AST
detection AUROC	92.2	94.4	91.8

types and quantities of testing samples can not fully verify the proposed model is reliable, which hinders the generalization ability of the model. Therefore, it is necessary to launch richer datasets with diverse scenarios and testing samples.

B. CONFLICT BETWEEN FAR AND MAR
In industrial applications there are intractably practical problems, FAR and MAR being a pair of contradictions. Correspondingly, the algorithm should be optimized to achieve a
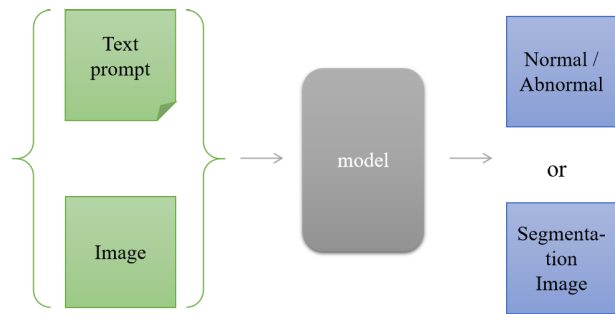
FIGURE 13. Future model pipeline. Input image and text prompt, and the model can output classification or defect segmentation.

reduction in both false alarm rate and missed alarm rate. Otherwise missed alarms can lead to the production of inferior products, which will cause commercial loss, whereas a high rate of false alarms can lead to increased costs for manual confirmation.

C. COMBINATION OF DATA DISTRIBUTION LEARNING AND DATA AUGMENTATION

Normalizing flow (NF)-based methods transform a simple distribution, such as a Gaussian distribution, into a more complex distribution by applying a series of invertible transformations. For unsupervised anomaly detection, NF is used to learn the distribution of normal samples [100]. There are other approaches [158] also based on the idea of learning data distribution. Considering the idea of learning data distribution in combination with data augmentation, NF can also be used to learn the distribution of artificially augmented defective samples. The joint learning of normal and artificial anomaly samples is beneficial to improve the generalization ability of the model.

D. FURTHER RESEARCH ON FOUNDATION MODEL

Representation-based methods apply a pre-trained model to extract image features for anomaly detection, which demonstrates the effectiveness of the pre-trained model. As data volume and model scale evolve, foundation model [159], [160] shows great potential as a member of pre-trained models. Foundation models are trained on massive amounts of data, which enables them to capture a broad range of patterns and relationships. By fine-tuning the model on specific tasks, it can quickly adapt to new domains and produce high-quality representation. The foundation model has striking strengths in representing ability and adaptation efficiency, and it has been started to be utilized in the field of computer vision [161], [162]. While the relevant research in the field of industrial anomaly detection still needs further exploration.

E. MULTIMODAL INDUSTRIAL ANOMALY DETECTION

Multimodal learning can facilitate deeper understanding by providing multiple perspectives and facilitating connections between different modalities. With the development of multimodal learning, models have shown great potential in dealing

with image and text modalities, like GPT-4 [163], CLIP [164], stable diffusion [165], SAM [166], OFA [167] and Unified-IO [168]. In future research on industrial anomaly detection, it is expected to accept image and text prompt inputs and produce specified results, such as normal/abnormal classification or defect segmentation, as shown in Fig. 13.

VII. CONCLUSION

Deep learning has inspired a surge of interest in the visual industrial anomaly detection problem in recent years, resulting in a wide range of creative solutions. We present a complete review of newly proposed methodologies for visual industrial anomaly detection in the literature in this study. We categorize the relevant approaches based on their fundamental principles and describe their assumptions, benefits, and drawbacks, which may be of interest to practitioners as well as academic researchers. We hope to assist academics in better understanding the common principles of visual industrial anomaly detection systems and identifying interesting research directions in this area. The unsupervised anomaly detection algorithm is still under continuous research and development, and we will continue to track the progress in the follow-up work.

REFERENCES

[1] X. Fang, Q. Luo, B. Zhou, C. Li, and L. Tian, "Research progress of automated visual surface defect detection for industrial metal planar materials," *Sensors*, vol. 20, no. 18, p. 5136, Sep. 2020.

[2] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *ELCVIA, Electron. Lett. Comput. Vis. Image Anal.*, vol. 7, no. 3, pp. 1–22, 2008.

[3] T. Zhang, Y. Liu, Y. Yang, X. Wang, and Y. Jin, "Review of surface defect detection based on machine vision," *J. Sci. Technol. Eng.*, vol. 20, no. 35, pp. 14366–14376, 2020.

[4] J. Cao, G. Yang, and X. Yang, "A pixel-level segmentation convolutional neural network based on deep feature fusion for surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[5] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, Mar. 2021.

[6] L. Wang and P. Koniusz, "Self-supervising action recognition by statistical moment and subspace descriptors," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 4324–4333.

[7] P. Koniusz, L. Wang, and A. Cherian, "Tensor representations for action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 648–665, Feb. 2022.

[8] L. Wang, D. Q. Huynh, and M. R. Mansour, "Loss switching fusion with similarity search for video classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 974–978.

[9] P.-H. Chen, Y.-C. Yang, and L.-M. Chang, "Automated bridge coating defect recognition using adaptive ellipse approach," *Autom. Construct.*, vol. 18, no. 5, pp. 632–643, Aug. 2009.

[10] S. Lee, L.-M. Chang, and M. Skibniewski, "Automated recognition of surface defects using digital color image processing," *Autom. Construct.*, vol. 15, no. 4, pp. 540–549, Jul. 2006.

[11] R. Manish, A. Venkatesh, and S. D. Ashok, "Machine vision based image processing techniques for surface finish and defect inspection in a grinding process," *Mater. Today, Proc.*, vol. 5, no. 5, pp. 12792–12802, 2018.

[12] M. Chu, R. Gong, S. Gao, and J. Zhao, "Steel surface defects recognition based on multi-type statistical features and enhanced twin support vector machine," *Chemometric Intell. Lab. Syst.*, vol. 171, pp. 140–150, Dec. 2017.

[13] H.-F. Ng, "Automatic thresholding for defect detection," *Pattern Recognit. Lett.*, vol. 27, no. 14, pp. 1644–1649, Oct. 2006.

[14] M. Aminzadeh and T. Kurfess, "Automatic thresholding for defect detection by background histogram mode extents," *J. Manuf. Syst.*, vol. 37, pp. 83–92, Oct. 2015.

[15] J. Kumar, S. P. Srivastava, R. S. Anand, P. Arvind, S. Bhardwaj, and A. Thakur, "GLCM and ANN based approach for classification of radiographics weld images," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2018, pp. 168–172.

[16] M. A. Shabir, M. U. Hassan, X. Yu, and J. Li, "Tyre defect detection based on GLCM and Gabor filter," in *Proc. 22nd Int. Multitopic Conf. (INMIC)*, Nov. 2019, pp. 1–6.

[17] S. Sadaghiyanfam, "Using gray-level-co-occurrence matrix and wavelet transform for textural fabric defect detection: A comparison study," in *Proc. Electr. Electron., Comput. Sci., Biomed. Eng. Meeting (EBBT)*, Apr. 2018, pp. 1–5.

[18] N. T. Deotale and T. K. Sarode, "Fabric defect detection adopting combined GLCM, Gabor wavelet features and random decision forest," *3D Res.*, vol. 10, no. 1, pp. 1–13, Mar. 2019.

[19] X. YongHua and W. Jin-Cong, "Study on the identification of the wood surface defects based on texture features," *Optik Int. J. Light Electron Opt.*, vol. 126, no. 19, pp. 2231–2235, Oct. 2015.

[20] S.-W. Yang, C.-S. Lin, S.-K. Lin, and H.-T. Chiang, "Automatic defect recognition of TFT array process using gray level co-occurrence matrix," *Optik*, vol. 125, no. 11, pp. 2671–2676, Jun. 2014.

[21] V. A. Sindagi and S. Srivastava, "Domain adaptation for automatic OLED panel defect detection using adaptive support vector data description," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 193–211, Apr. 2017.

[22] K. Song and Y. Yan, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl. Surf. Sci.*, vol. 285, pp. 858–864, Nov. 2013.

[23] S. Fekri-Ershad and F. Tajeripour, "Multi-resolution and noise-resistant surface defect detection approach using new version of local binary patterns," *Appl. Artif. Intell.*, vol. 31, nos. 5–6, pp. 395–410, Jul. 2017.

[24] Z. Liu, L. Yan, C. Li, Y. Dong, and G. Gao, "Fabric defect detection based on sparse representation of main local binary pattern," *Int. J. Clothing Sci. Technol.*, vol. 29, no. 3, pp. 282–293, Jun. 2017.

[25] S. Li, D. Li, and W. Yuan, "Wood defect classification based on two-dimensional histogram constituted by LBP and local binary differential excitation pattern," *IEEE Access*, vol. 7, pp. 145829–145842, 2019.

[26] K. Yan, Q. Dong, T. Sun, M. Zhang, and S. Zhang, "Weld defect detection based on completed local ternary patterns," in *Proc. Int. Conf. Video Image Process.*, Dec. 2017, pp. 6–14.

[27] A. Chondronasios, I. Popov, and I. Jordanov, "Feature selection for surface defect classification of extruded aluminum profiles," *Int. J. Adv. Manuf. Technol.*, vol. 83, nos. 1–4, pp. 33–41, Mar. 2016.

[28] R. Ma, S. Deng, H. Sun, and Y. Qi, "An algorithm for fabric defect detection based on adaptive Canny operator," in *Proc. Int. Conf. Intell. Comput., Autom. Syst. (ICICAS)*, Dec. 2019, pp. 475–481.

[29] Y. Zhang, T. Li, and Q. Li, "Defect detection for tire laser shearography image using curvelet transform based edge detector," *Opt. Laser Technol.*, vol. 47, pp. 64–71, Apr. 2013.

[30] X. Hou and H. Liu, "Welding image edge detection and identification research based on Canny operator," in *Proc. Int. Conf. Comput. Sci. Service Syst.*, Aug. 2012, pp. 250–253.

[31] M. Halim, "Automatic laser welding defect detection and classification using sobel-contour shape detection," *J. Telecommun., Electron. Comput. Eng. (JTEC)*, vol. 8, no. 6, pp. 157–160, 2016.

[32] C. Tastimur, M. Karaköse, E. Akin, and I. Aydin, "Rail defect detection with real time image processing technique," in *Proc. IEEE 14th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2016, pp. 411–415.

[33] J. Zhang, Z. Guo, T. Jiao, and M. Wang, "Defect detection of aluminum alloy wheels in radiography images using adaptive threshold and morphological reconstruction," *Appl. Sci.*, vol. 8, no. 12, p. 2365, Nov. 2018.

[34] J. Hu, Z. He, G. Weng, L. Sun, B. Zuo, and C. Wang, "Detection of chemical fabric defects on the basis of morphological processing," *J. Textile Inst.*, vol. 107, no. 2, pp. 233–241, Feb. 2016.

[35] V. Jayashree and S. Subbaramn, "Hybrid approach using correlation and morphological approaches for GFDD of plain weave fabric," in *Proc. IEEE Control Syst. Graduate Res. Colloq.*, Jul. 2012, pp. 197–202.

[36] A. Rebhi, S. Abid, and F. Fnaiech, "Fabric defect detection using local homogeneity and morphological image processing," in *Proc. Int. Image Process., Appl. Syst. (IPAS)*, Nov. 2016, pp. 1–5.

[37] T.-C. Su and M.-D. Yang, "Application of morphological segmentation to leaking defect detection in sewer pipelines," *Sensors*, vol. 14, no. 5, pp. 8686–8704, May 2014.

[38] I. Pastor-López, B. Sanz, J. G. D. L. Puerta, and P. G. Bringas, "Surface defect modelling using co-occurrence matrix and fast Fourier transformation," in *Proc. Int. Conf. Hybrid Artif. Intell. Syst.*, 2019, pp. 745–757.

[39] X. Bai, Y. Fang, W. Lin, L. Wang, and B.-F. Ju, "Saliency-based defect detection in industrial images by using phase spectrum," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2135–2145, Nov. 2014.

[40] P. Kumaresan, "Defect detection in texture by Fourier analysis approach," *Int. J. Eng., Sci. Math.*, vol. 6, no. 3, pp. 124–130, 2017.

[41] S. Ajithaprasad, R. Velpula, and R. Gannavarpu, "Defect detection using windowed Fourier spectrum analysis in diffraction phase microscopy," *J. Phys. Commun.*, vol. 3, no. 2, Feb. 2019, Art. no. 025006.

[42] G.-H. Hu, "Automated defect detection in textured surfaces using optimal elliptical Gabor filters," *Optik*, vol. 126, no. 14, pp. 1331–1340, Jul. 2015.

[43] J.-W. Wang, W.-Y. Chen, and J.-S. Lee, "Singular value decomposition combined with wavelet transform for lcd defect detection," *Electron. Lett.*, vol. 48, no. 5, pp. 266–267, 2012.

[44] C. Yang, P. Liu, G. Yin, H. Jiang, and X. Li, "Defect detection in magnetic tile images based on stationary wavelet transform," *NDT E Int.*, vol. 83, pp. 78–87, Oct. 2016.

[45] J. Liu, G. Xu, L. Ren, Z. Qian, and L. Ren, "Defect intelligent identification in resistance spot welding ultrasonic detection based on wavelet packet and neural network," *Int. J. Adv. Manuf. Technol.*, vol. 90, nos. 9–12, pp. 2581–2588, Jun. 2017.

[46] P. Li, H. Zhang, J. Jing, R. Li, and J. Zhao, "Fabric defect detection based on multi-scale wavelet transform and Gaussian mixture model method," *J. Textile Inst.*, vol. 106, no. 6, pp. 587–592, Jun. 2015.

[47] J. Jing, P. Yang, P. Li, and X. Kang, "Supervised defect detection on textile fabrics via optimal Gabor filter," *J. Ind. Textiles*, vol. 44, no. 1, pp. 40–57, Jul. 2014.

[48] L. Tong, W. K. Wong, and C. K. Kwong, "Differential evolution-based optimal Gabor filter model for fabric inspection," *Neurocomputing*, vol. 173, pp. 1386–1401, Jan. 2016.

[49] X. Yang, D. Qi, and X. Li, "Multi-scale edge detection of wood defect images based on the dyadic wavelet transform," in *Proc. Int. Conf. Mach. Vis. Hum. Mach. Interface*, 2010, pp. 120–123.

[50] V. R. Vijaykumar and S. Sangamithirai, "Rail defect detection using Gabor filters with texture analysis," in *Proc. 3rd Int. Conf. Signal Process., Commun. Netw. (ICSCN)*, Mar. 2015, pp. 1–6.

[51] Y. Li, W. Zhao, and J. Pan, "Deformable patterned fabric defect detection with Fisher criterion-based deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1256–1264, Apr. 2017.

[52] Y. Wang, M. Liu, P. Zheng, H. Yang, and J. Zou, "A smart surface inspection system using faster R-CNN in cloud-edge computing environment," *Adv. Eng. Informat.*, vol. 43, Jan. 2020, Art. no. 101037.

[53] Y. T. Liu, Y. N. Yang, W. Chao, X. Y. Xu, and T. Zhang, "Research on surface defect detection based on semantic segmentation," *DEStech Trans. Comput. Sci. Eng.*, 2019.

[54] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Ann.*, vol. 65, no. 1, pp. 417–420, 2016.

[55] Y. Li and C. Zhang, "Automated vision system for fabric defect inspection using Gabor filters and PCNN," *SpringerPlus*, vol. 5, no. 1, pp. 1–12, Dec. 2016.

[56] S. S. Kumar, D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. Starr, "Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks," *Autom. Construct.*, vol. 91, pp. 273–283, Jul. 2018.

[57] D. Brackenbury, I. Brilakis, and B. DeJong, "Automated defect detection for masonry arch bridges," in *Proc. Int. Conf. Smart Infrastruct. Construct. (ICSIC)*, Jan. 2019, pp. 3–9.

[58] C. Sacco et al., "Automated fiber placement defects: Automated inspection and characterization," in *Proc. SAMPE Conf. Exhib.*, 2018.

[59] Y. Yang, L. Pan, J. Ma, R. Yang, Y. Zhu, Y. Yang, and L. Zhang, "A high-performance deep learning algorithm for the automated optical inspection of laser welding," *Appl. Sci.*, vol. 10, no. 3, p. 933, Jan. 2020.

[60] Y. Liu, J. Geng, Z. Su, W. Zhang, and J. Li, "Real-time classification of steel strip surface defects based on deep CNNs," in *Proc. Chin. Intell. Syst. Conf.*, 2019, pp. 257–266.

[61] L. Song, W. Lin, Y. Yang, X. Zhu, Q. Guo, and J. Xi, "Weak micro-scratch detection based on deep convolutional neural network," *IEEE Access*, vol. 7, pp. 27547–27554, 2019.

[62] Y. Liu, K. Xu, and J. Xu, "Periodic surface defect detection in steel plates based on deep learning," *Appl. Sci.*, vol. 9, no. 15, p. 3127, Aug. 2019.

[63] R. F. Pereira, C. M. S. Medeiros, and P. P. R. Filho, "Goat leather quality classification using computer vision and machine learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[64] S. Tural and R. Samet, "Automated defect detection on surface of Militarz cartridges," *J. Mod. Technol. Eng.*, vol. 4, pp. 178–189, Jan. 2019.

[65] M. P. Samy, S. Foong, G. S. Soh, and K. S. Yeo, "Automatic optical & laser-based defect detection and classification in brick masonry walls," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2016, pp. 3521–3524.

[66] F. López, J.-M. Valiente, and J.-M. Prats, "Surface grading using soft colour-texture descriptors," in *Proc. Iberoamerican Congr. Pattern Recognit.*, 2005, pp. 13–23.

[67] F. López, J. M. Valiente, R. Baldrich, and M. Vanrell, "Fast surface grading using color statistics in the cie lab space," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2005, pp. 666–673.

[68] F. Pernkopf, "Detection of surface defects on raw steel blocks using Bayesian network classifiers," *Pattern Anal. Appl.*, vol. 7, no. 3, pp. 333–342, Dec. 2004.

[69] C. Mandriota, M. Nitti, N. Ancona, E. Stella, and A. Distante, "Filter-based feature selection for rail defect detection," *Mach. Vis. Appl.*, vol. 15, no. 4, pp. 179–185, Oct. 2004.

[70] K. Wiltschi, A. Pinz, and T. Lindeberg, "An automatic assessment scheme for steel quality inspection," *Mach. Vis. Appl.*, vol. 12, no. 3, pp. 113–128, Oct. 2000.

[71] C.-H. Chan and G. K. H. Pang, "Fabric defect detection by Fourier analysis," *IEEE Trans. Ind. Appl.*, vol. 36, no. 5, pp. 1267–1276, Sep./Oct. 2000.

[72] A. Latif-Amet, A. Ertüzün, and A. Erçil, "An efficient method for texture defect detection: Sub-band domain co-occurrence matrices," *Image Vis. Comput.*, vol. 18, nos. 6–7, pp. 543–553, May 2000.

[73] R. Wang, K. Nie, Y. Chang, X. Gong, T. Wang, Y. Yang, and B. Long, "Deep learning for anomaly detection," in *Proc. KDD*, 2020, pp. 3569–3570.

[74] Y. Gao, X. Li, X. V. Wang, L. Wang, and L. Gao, "A review on recent advances in vision-based defect recognition towards industrial intelligence," *J. Manuf. Syst.*, vol. 62, pp. 753–766, Jan. 2022.

[75] X. Zheng, S. Zheng, Y. Kong, and J. Chen, "Recent advances in surface defect inspection of industrial products using deep learning techniques," *Int. J. Adv. Manuf. Technol.*, vol. 113, nos. 1–2, pp. 35–58, Mar. 2021.

[76] J. Yang, R. Xu, Z. Qi, and Y. Shi, "Visual anomaly detection for images: A systematic survey," *Proc. Comput. Sci.*, vol. 199, pp. 471–478, Jan. 2022.

[77] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C. M. Oddo, and P. Dario, "Visual-based defect detection and classification approaches for industrial applications—A SURVEY," *Sensors*, vol. 20, no. 5, p. 1459, Mar. 2020.

[78] S. Qi, J. Yang, and Z. Zhong, "A review on industrial surface defect detection based on deep learning technology," in *Proc. 3rd Int. Conf. Mach. Learn. Mach. Intell.*, Sep. 2020, pp. 24–30.

[79] A. Rasheed, B. Zafar, A. Rasheed, N. Ali, M. Sajid, S. H. Dar, U. Habib, T. Shehryar, and M. T. Mahmood, "Fabric defect detection using computer vision techniques: A comprehensive review," *Math. Problems Eng.*, vol. 2020, pp. 1–24, Nov. 2020.

[80] P. M. Bhatt, R. K. Malhan, P. Rajendran, B. C. Shah, S. Thakar, Y. J. Yoon, and S. K. Gupta, "Image-based surface defect detection using deep learning: A review," *J. Comput. Inf. Sci. Eng.*, vol. 21, no. 4, pp. 1–23, Aug. 2021.

[81] M. Salehi, H. Mirzaei, D. Hendrycks, Y. Li, M. Hossein Rohban, and M. Sabokrou, "A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges," 2021, *arXiv:2110.14051*.

[82] X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "GAN-based anomaly detection: A review," *Neurocomputing*, vol. 493, pp. 497–535, Jul. 2022.

[83] N. Goernitz, M. Kloft, K. Rieck, and U. Brefeld, "Toward supervised anomaly detection," *J. Artif. Intell. Res.*, vol. 46, pp. 235–262, Feb. 2013.

[84] Y. Kawachi, Y. Koizumi, and N. Harada, "Complementary set variational autoencoder for supervised anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2366–2370.

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

[86] S. Venkataramanan, K.-C. Peng, R. V. Singh, and A. Mahalanobis, "Attention guided anomaly localization in images," in *Proc. Eur. Conf. Comput. Vis.*, Nov. 2020, pp. 485–503.

[87] Y. Liu, C. Zhuang, and F. Lu, "Unsupervised two-stage anomaly detection," 2021, *arXiv:2103.11671*.

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

[89] M. Zolfaghari and H. Sajedi, "Unsupervised anomaly detection with an enhanced teacher for student-teacher feature pyramid matching," in *Proc. 27th Int. Comput. Conf., Comput. Soc. Iran (CSICC)*, Feb. 2022, pp. 1–4.

[90] S. Yamada, S. Kamiya, and K. Hotta, "Reconstructed student-teacher and discriminative networks for anomaly detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 2725–2732.

[91] 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. 9727–9736.

[92] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Asymmetric student-teacher networks for industrial anomaly detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 2591–2601.

[93] F. V. Massoli, F. Falchi, A. Kantarci, Ş. Akti, H. K. Ekenel, and G. Amato, "MOCCA: Multi-layer one-class ClassificAtion for anomaly detection," 2020, *arXiv:2012.12111*.

[94] Y. Liang, J. Zhang, S. Zhao, R. Wu, Y. Liu, and S. Pan, "Omni-frequency channel-selection representations for unsupervised anomaly detection," 2022, *arXiv:2203.00259*.

[95] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, "VT-ADL: A vision transformer network for image anomaly detection and localization," in *Proc. IEEE 30th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2021, pp. 01–06.

[96] N. Li, K. Jiang, Z. Ma, X. Wei, X. Hong, and Y. Gong, "Anomaly detection via self-organizing map," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 974–978.

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

[98] 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.

[99] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Fully convolutional cross-scale-flows for image-based defect detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1829–1838.

[100] J. Yu, J. Zheng, X. Wang, W. Li, Y. Wu, R. Zhao, and L. Wu, "FastFlow: Unsupervised anomaly detection and localization via 2D normalizing flows," 2021, *arXiv:2111.07677*.

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

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

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

[104] S. Wang, L. Wu, L. Cui, and Y. Shen, "Glancing at the patch: Anomaly localization with global and local feature comparison," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 254–263.

[105] Y. Zheng, X. Wang, R. Deng, T. Bao, R. Zhao, and L. Wu, "Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization," in *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, Jul. 2022, pp. 1–6.

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

[107] J. Yi and S. Yoon, "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 375–390.

[108] K. Sohn, C. L. Li, J. Yoon, M. Jin, and T. Pfister, "Learning and evaluating representations for deep one-class classification," 2020, *arXiv:2011.02578*.

[109] J.-H. Kim, D.-H. Kim, S. Yi, and T. Lee, "Semi-orthogonal embedding for efficient unsupervised anomaly segmentation," 2021, *arXiv:2105.14737*.

[110] S. Lee, S. Lee, and B. C. Song, "CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization," *IEEE Access*, vol. 10, pp. 78446–78454, 2022.

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

[112] H. M. Schlüter, J. Tan, B. Hou, and B. Kainz, "Natural synthetic anomalies for self-supervised anomaly detection and localization," 2021, *arXiv:2109.15222*.

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

[114] N. Ristea, N. Madan, R. T. Ionescu, K. Nasrollahi, F. S. Khan, T. B. Moeslund, and M. Shah, "Self-supervised predictive convolutional attentive block for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13566–13576.

[115] J. Song, K. Kong, Y.-I. Park, S.-G. Kim, and S.-J. Kang, "AnoSeg: Anomaly segmentation network using self-supervised learning," 2021, *arXiv:2110.03396*.

[116] Y. LeCun, "Generalization and network design strategies," *Connectionism Perspective*, vol. 19, pp. 143–155, Jun. 1989.

[117] M. Rudolph, B. Wandt, and B. Rosenhahn, "Structuring autoencoders," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 615–623.

[118] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[119] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[120] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 37, 2015, pp. 1530–1538.

[121] X. Wang, K. Wang, and S. Lian, "A survey on face data augmentation for the training of deep neural networks," *Neural Comput. Appl.*, vol. 32, pp. 15503–15531, Mar. 2020.

[122] N. Ishida, Y. Nagatsu, and H. Hashimoto, "Unsupervised anomaly detection based on data augmentation and mixing," in *Proc. IECON 46th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2020, pp. 529–533.

[123] Y. Chen, Y. Tian, G. Pang, and G. Carneiro, "Deep one-class classification via interpolated Gaussian descriptor," in *AAAI*, 2022, pp. 383–392.

[124] P. Liznerski, L. Ruff, R. A. Vandermeulen, B. J. Franks, M. Kloft, and K. R. Müller, "Explainable deep one-class classification," *arXiv:2007.01760*, 2020.

[125] T. Reiss and Y. Hoshen, "Mean-shifted contrastive loss for anomaly detection," 2021, *arXiv:2106.03844*.

[126] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.

[127] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. ICML*, 2018, pp. 4393–4402.

[128] J. Pirnay and K. Chai, "Inpainting transformer for anomaly detection," in *Image Analysis and Processing—ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part II*. Cham, Switzerland: Springer, 2022.

[129] A. M. Kamoona, A. K. Gostar, A. Bab-Hadiashar, and R. Hoseinnezhad, "Anomaly detection of defect using energy of point pattern features within random finite set framework," 2021, *arXiv:2108.12159*.

[130] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[131] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 372–380.

[132] X. Yao, C. Zhang, and R. Li, "Explicit boundary guided semi-push-pull contrastive learning for better anomaly detection," 2022, *arXiv:2207.01463*.

[133] S. Deitsch, V. Christlein, S. Berger, C. Buerhop-Lutz, A. Maier, F. Gallwitz, and C. Riess, "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Sol. Energy*, vol. 185, pp. 455–468, Jun. 2019.

[134] C. Buerhop-Lutz, S. Deitsch, A. Maier, F. Gallwitz, S. Berger, B. Doll, J. Hauch, C. Camus, and C. J. Brabec, "A benchmark for visual identification of defective solar cells in electroluminescence imagery," in *Proc. Eur. PV Sol. Energy Conf. Exhib. (EU PVSEC)*, 2018, pp. 1287–1289.

[135] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steland, A. Maier, F. Gallwitz, and C. Riess, "Segmentation of photovoltaic module cells in uncalibrated electroluminescence images," *Mach. Vis. Appl.*, vol. 32, no. 4, pp. 32–84, 2021.

[136] J. Silvestre-Blanes, T. Albero-Albero, I. Miralles, R. Pérez-Llorens, and J. Moreno, "A public fabric database for defect detection methods and results," *Autex Res. J.*, vol. 19, no. 4, pp. 363–374, Dec. 2019.

[137] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14897–14907.

[138] G. Pang, C. Ding, C. Shen, and A. van den Hengel, "Explainable deep few-shot anomaly detection with deviation networks," 2021, *arXiv:2108.00462*.

[139] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[140] J. Tack, S. Mo, J. Jeong, and J. Shin, "CSI: Novelty detection via contrastive learning on distributionally shifted instances," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 11839–11852.

[141] W. Liu, W. Luo, Z. Li, P. Zhao, and S. Gao, "Margin learning embedded prediction for video anomaly detection with a few anomalies," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3023–3030.

[142] C. Ding, G. Pang, and C. Shen, "Catching both gray and black swans: Open-set supervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7378–7388.

[143] Y. Huang, C. Qiu, and K. Yuan, "Surface defect saliency of magnetic tile," *Vis. Comput.*, vol. 36, no. 1, pp. 85–96, Jan. 2020.

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

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

[146] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang, "Deep structured energy based models for anomaly detection," in *Proc. ICML*, 2016, pp. 1100–1109.

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

[148] H. Cheng, H. Liu, F. Gao, and Z. Chen, "ADGAN: A scalable GAN-based architecture for image anomaly detection," in *Proc. IEEE 4th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, vol. 1, Jun. 2020, pp. 987–993.

[149] J. Andrews et al., "Transfer representation-learning for anomaly detection," *JMLR*, 2016.

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

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

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

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

[154] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[155] 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.

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

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

[158] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, *arXiv:1511.05644*.

[159] R. Bommasani, "On the opportunities and risks of foundation models," 2021, *arXiv:2108.07258*.

[160] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 15979–15988.

[161] T. Xi, "UFO: Unified feature optimization," in *Proc. Eur. Conf. Comput. Vis.* Tel Aviv, Israel: Springer, Oct. 2022, pp. 472–488.

[162] C. Riquelme, J. Puigcerver, B. Mustafa, M. Neumann, R. Jenatton, A. S. Pinto, D. Keysers, and N. Houlsby, "Scaling vision with sparse mixture of experts," in *Proc. Adv. Neural Inf. Process. Syst., Annu. Conf. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, Dec. 2021, pp. 8583–8595.

[163] OpenAI, "GPT-4 technical report," 2023, *arXiv:2303.08774*.

[164] A. Radford, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8748–8763.

[165] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10674–10685.

[166] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023, *arXiv:2304.02643*.

[167] P. Wang, A. Yang, R. Men, J. Lin, S. Bai, Z. Li, J. Ma, C. Zhou, J. Zhou, and H. Yang, "OFA: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 23318–23340.

[168] J. Lu, C. Clark, R. Zellers, R. Mottaghi, and A. Kembhavi, "Unified-IO: A unified model for vision, language, and multi-modal tasks," 2022, *arXiv:2206.08916*.

YAJIE CUI received the B.S. degree in computer science and technology from Northeastern University, Shenyang, China, in 2018, and the M.S. degree in computer science and technology from Tianjin University, Tianjin, China, in 2021.

Since 2021, she has been an Algorithm Engineer in deep learning with the AI Innovation and Application Center, China Unicom Digital Technology Company Ltd., Beijing, China. Her research interests include anomaly detection in computer vision, machine learning, and deep learning.

ZHAOXIANG LIU received the B.S. and Ph.D. degrees from the College of Information and Electrical Engineering, China Agricultural University, Beijing, China, in 2006 and 2011, respectively.

He joined VIA Technologies Inc., Beijing, in 2011. From 2012 to 2016, he was a Senior Researcher with the Central Research Institute, Huawei Technologies, Beijing. He was a Senior Manager of CloudMinds Technologies Inc., Beijing, from 2016 to 2019. Since 2019, he has been the Director of the AI Innovation and Application Center, China Unicom Digital Technology Company Ltd., Beijing. His current research interests include computer vision, deep learning, robotics, and human–computer interaction.

SHIGUO LIAN (Member, IEEE) received the Ph.D. degree from the Nanjing University of Science and Technology, China.

He was a Research Assistant with the City University of Hong Kong, Hong Kong, in 2004. From 2005 to 2010, he was a Research Scientist with France Telecom Research and Development Beijing, Beijing, China. He was a Senior Research Scientist and the Technical Director of the Huawei Central Research Institute, Beijing, from 2010 to 2016. He was the Senior Director of CloudMinds Technologies Inc., Beijing, from 2016 to 2019. Since 2019, he has been the Chief AI Scientist of China Unicom Digital Technology Company Ltd., and the General Manager of the AI Innovation and Application Center, China Unicom Digital Technology Company Ltd., Beijing. He has authored or coedited over ten books. He holds over 50 patents. He has authored over 80 refereed international journal articles covering the topics of artificial intelligence, multimedia communication, and human–computer interface.

Dr. Lian is on the editor board of several refereed international journals.

• • •