Machine Intelligence Research
www.mi-research.net

21(1), February 2024, 104-135
DOI: 10.1007/s11633-023-1459-z

Deep Industrial Image Anomaly Detection: A Survey

Jiaqi Liu[1*] Guoyang Xie[1,2*] Jinbao Wang[1*] Shangnian Li[1]
Chengjie Wang[3] Feng Zheng[1] Yaochu Jin[2,4]

[1]Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen 518055, China

[2]NICE Group, University of Surrey, Guildford GU2 7YX, UK

[3]Youtu Lab, Tencent, Shanghai 200233, China

[4]NICE Group, Bielefeld University, Bielefeld 33619, Germany

Abstract: The recent rapid development of deep learning has laid a milestone in industrial image anomaly detection (IAD). In this paper, we provide a comprehensive review of deep learning-based image anomaly detection techniques, from the perspectives of neural network architectures, levels of supervision, loss functions, metrics and datasets. In addition, we extract the promising setting from industrial manufacturing and review the current IAD approaches under our proposed setting. Moreover, we highlight several opening challenges for image anomaly detection. The merits and downsides of representative network architectures under varying supervision are discussed. Finally, we summarize the research findings and point out future research directions. More resources are available at https://github.com/M-3LAB/awesome-industrial-anomaly-detection.

Keywords: Image anomaly detection, defect detection, industrial manufacturing, deep learning, computer vision.

Citation: J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, Y. Jin. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, vol.21, no.1, pp.104–135, 2024. http://doi.org/10.1007/s11633-023-1459-z

1 Introduction

We review the recent advances of deep learning-based image anomaly detection since the rapid development of deep learning can bring the capabilities of image anomaly detection into the factory floor. In modern manufacturing, image anomaly detection (IAD) is always performed at the end of the manufacturing process and tries to identify product defects. The price of a product is significantly affected by the defect′s severity. In addition, if the flaw reaches a certain threshold, the product will be discarded. Historically, the majority of anomaly detection tasks are performed by humans, which suffers from the following many disadvantages:

1) It is impossible to avoid human fatigue, resulting in a false positive phenomenon (i.e., the ground truth is abnormal, while the human′s judgment is normal).

2) Long and intensive work on anomaly detection may cause health problems, such as visual impairment.

3) Locating anomalies requires a significant number of employees, raising operational costs.

Thus, the goal of IAD algorithms is to reduce human labour and improve productivity and product quality. Before deep learning, the performance of IAD could not fulfil the demands of industrial manufacturing. Nowadays, the deep learning method has received good results, and most of these methods are more than 97% accurate. Still, IAD has many problems when it comes to real-world use. To comprehensively explore the effectiveness and applicable scenarios of the current methods, more careful analysis of IAD we conduct in this survey is necessary and significant.

Table 1 demonstrates clearly the merits of our survey in terms of dataset, metric, neural network architecture, levels of supervision and promising setting for industrial manufacturing. As a representative review that focuses more on traditional methods, Czimmermann et al.[1] have less discussion of deep learning methods, while our survey discusses deep learning in more depth. Firstly, our study uses twice as many IAD datasets as Tao et al.[2] Secondly, we analyze the performance of IAD using the most comprehensive image level and pixel level metrics. Nevertheless, Cui et al.[3] and Tao et al.[2] only employ image level metrics, neglecting the anomalies localization performance of IAD. Thirdly, our study develops a taxonomy based on the design of neural network architecture with varying degrees of supervision. Finally, to bridge the gap between academic research and real-world industry needs, we review the current IAD algorithms under industrial manufacturing settings.

As an emerging field, research on IAD must fully consider industrial manufacturing requirements. The follow-

Review
Manuscript received on March 22, 2023; accepted on May 26, 2023
Recommended by Associate Editor Mao-Guo Gong
Colored figures are available in the online version at https://link.springer.com/journal/11633
* These authors contribute equally to this work

Table 1 Related surveys and ours for IAD

Content	Czimmermann et al.[1]	Tao et al.[2]	Cui et al.[3]	Ours
IAD dataset	–	9	7	**20**
IAD metric	–	3	1	**6**
Neural network architecture	×	√	√	√
Levels of supervision	√	×	×	√
Industrial manufacturing setting	×	×	×	√

ing is a summary of the challenging issues that need to be investigated:

1) IAD dataset should be gathered from actual manufacturing lines, not labs. The public cannot access the real-world anomalous dataset due to privacy concerns. The majority of open-source IAD datasets generate anomalies from anomaly-free products. In other words, the abnormalities from open-source IAD datasets may not occur in actual production lines, which makes deploying IADs in industrial manufacturing very challenging.

2) It is challenging to enable the creation of a unified IAD model in the absence of multiple domain IAD datasets. Recently, You et al.[4] propose a unified IAD model for multiple class objects. However, they disregard the notion that commodities produced in the same plant should be of the same sort. For example, an automaker manufactures several types of workpieces but does not produce fruit. Current popular IAD datasets, like MVTec AD[5] and MVTec LOCO[6], consist of numerous classes but not multiple domains. To simulate a realistic manufacturing process, we must create a new IAD dataset collected from multiple domains.

3) It is urgent to set up a uniform assessment for the image-level and pixel level of IAD performance. The majority of IAD metrics shrink the anomalous mask (ground truth) into the size of feature map for evaluation, which inevitably reduces the precision of assessment. Moreover, we discover that certain IAD methods perform well on image AUROC but poorly on pixel AP, or vice versa. Therefore, it is essential to develop a uniform metric for assessment IAD performance at both image and pixel level.

4) We should design a more efficient loss function that can leverage both the guidance of labelled data and the exploration of unlabelled data. In realistic manufacturing scenario, limited number of anomalous samples are available. However, most of unsupervised IAD methods outperform semi-supervised IAD methods. By observing the failure of semi-supervised IAD, we would call for more attention to the feature extraction and loss function, which can leverage both the guidance from labels efficiently and the exploration from the unlabeled data. Regarding the key problem mentioned above, improving feature extraction from abnormal samples and redesigning deviation loss function can fully use labelled anomalies and diverge the feature space of abnormal samples from those of normal samples.

The paper categorizes various methods into several paradigms, and clearly analyzes the advantages and disadvantages of various paradigms. It allows the reader to understand the state-of-the-art quickly and provides a reliable guide for selecting the required algorithm for practical applications. More importantly, we have analyzed the disadvantages of different paradigms and the current main challenges. Subsequent researchers can quickly find directions to push the field forward.

1.1 Contributions

The main contributions of this survey can be summarized as following:

1) We provide an in-depth review of image anomaly detection by considering the design of neural network architecture with varying degrees of supervision.

2) It provides a comprehensive review of the current IAD algorithms in different settings to bridge the gap between the academic research and real-world industrial manufacturing.

3) It summarizes the main issues and potential challenges in IAD, which outlines the underlying research directions for future works.

The rest of this paper is organized as Fig. 1. In Sections 2 and 3, we review IAD on the basis of the neural network architecture with different levels of supervision. Next, we review the recent advances of IAD under our proposed setting from industrial manufacturing in Section 4. We describe the popular dataset in Section 5 and take a retrospective view of the metrics function in Section 5. Then, we provide an analysis of the performance of current IAD methods on various datasets in Section 6. Finally, we provide future research directions for IAD in Section 7.

2 Unsupervised anomaly detection

The majority of current research focuses on unsupervised anomaly detection, based on the assumption that the collection of abnormal samples incurs massive human and financial costs. This indicates that only normal samples are included in the training set, whereas both abnormal and normal samples are included in the test set. Anomaly detection in industrial images is a subset of problems with out-of-distribution (OOD). Before the rise of deep learning, differential detection and filtering were frequently used to detect anomalies in industrial images. Following the release of the MVTec AD[5], methods for anomaly detection in industrial images can be divided into two categories: feature-embedding and reconstructed-based. Currently, more AD techniques are based on fea-

ture embedding.

2.1 Feature embedding based methods

2.1.1 Teacher-student architecture

The performance of these methods is outstanding, but they depend on pre-trained models such as ResNet[7] VGG[8] and EfficientNet[9]. The selection of the ideal teacher model is crucial. This type of instructional strategy is summarized in Table 2. The structure of the network and the method of distillation are the primary distinctions between various techniques.

The teacher-student network architecture depicted in Fig. 2 is the most standard technique for detecting industrial image anomalies. This method typically selects a partial layer of a backbone network pre-trained on a large-scale dataset as a fixed-parameter teacher model. During training, the teacher model imparts to the student model the knowledge of extracting normal sample features. During inference, the characteristics of normal images extracted from the test set by the teacher network and the student network are comparable, whereas the characteristics of abnormal images extracted from the test set are quite distinct. By comparing the feature maps generated by the two networks, it is possible to generate anomaly score maps with the same size. Then, by enlarging the anomaly score map to the same proportion as the input image, we can obtain the anomaly scores of various input image locations. On the justification of this model, it is possible to determine whether the test image is abnormal.

Bergmann et al.[10] are the first to use teacher-student architecture for anomaly detection. The model is straightforward and effective, significantly outperforming other benchmark methods. While STPM[12] and MKD[11] both use multi-scale features under different network layers for distillation, they do so in different ways. In this instance, the normal sample features extracted by the student network are more similar to those extracted by the teacher network, whereas the abnormal sample features are more dissimilar. In addition, MKD finds that the lighter student network structure performs better than the student network structure identical to that of the teacher network. Based on STPM, RSTPM[13, 17] adds a pair of teacher-student networks. During reasoning, the new teacher network is placed behind the original teacher-stu-

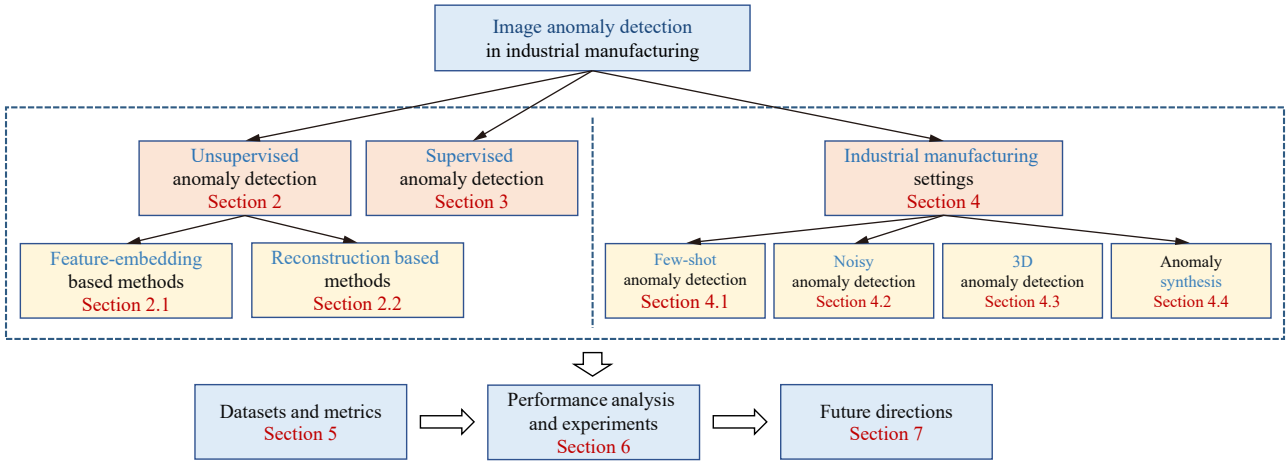

Fig. 1 Framework of this survey

Table 2 A summary of teacher-student methods regarding loss function, pre-trained model, and highlights

Method	Loss function	Pre-trained	Highlights
Uninformed students[10]	L_2, compactness	ResNet	The paper designs a basic approach to anomaly detection problems using a teacher-student model.
MKD[11]	L_2	VGG	The paper uses multi-scale features and lighter networks for distillation.
STPM[12]	L_2	ResNet	The paper uses multi-scale features under different network layers for distillation.
STFPM[13]	L_2	ResNet	The paper adds another teacher-student pair to get different feature reconstruction results.
RD4AD[14]	Cosine similarity	ResNet	The paper designs the teacher-student model of reverse distillation in a similar way to reconstruction.
IKD[15]	Context similarity	ResNet	The paper adds context similarity loss and adaptive hard sample mining module to prevent overfitting.
AST[16]	L_2, log-likelihood	EfficientNet	The paper uses an asymmetric teacher-student network to make the representation of anomaly more different.

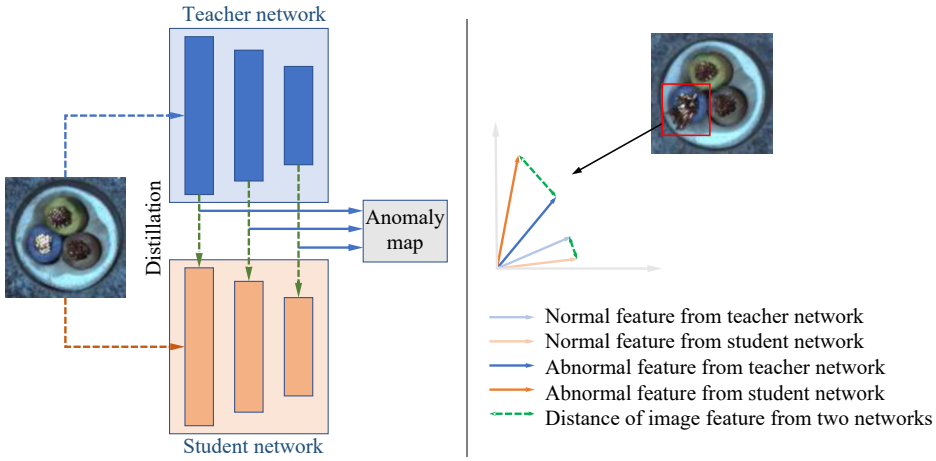

Fig. 2 Architecture of teacher-student models

dent network and is responsible for recreating the features. When anomalous images are presented, the student network typically reconstructs normal features that can be distinguished from those of the teacher network. RSTPM also includes a mechanism for transferring features from the teacher network to the student network in order to facilitate feature reconstruction. RD4AD[14] and RSTPM share certain similarities in their learning. RSTPM employs two pairs of teacher-student networks for feature reconstruction, whereas RD4AD only employs one pair of teacher-student networks. RD4AD proposes a multi-scale feature fusion (MFF) block and one-class bottleneck (OCB) to form an embedding, which is used to eliminate redundant features at multiple scales so that a single pair of teacher-student networks can perform feature reconstruction effectively. The abnormal image features extracted by the teacher-student network of RD4AD differ significantly during inference. AST[16] concludes that the abnormal image features extracted by the teacher-student model with the same structure are significantly similar, so they propose an asymmetric teacher-student architecture to address this issue. AST also introduces a normalized flow to avoid this problem and prevent estimation bias caused by the inconsistency of the two network structures. Previous teacher-student architecture anomaly detection methods suffer from overfitting as a result of inconsistency between neural network capacity and knowledge amount. By incorporating the context similarity loss (CSL) and adaptive hard sample mining (AHSM) modules, informative knowledge distillation (IKD)[15] hopes to reduce overfitting. CSL can assist the student network in comprehending the structure of a context-containing data manifold. The AHSM can concentrate on difficult samples containing a lot of information.

2.1.2 One-class classification

One-class classification techniques rely more heavily on abnormal samples. If the generated abnormal samples are of poor quality, the method's performance will be severely compromised. As demonstrated in Table 3, with the exception of MemSeg[18], the training of other methods relies on SVDD and cross-entropy loss. Consequently, the performance of the vast majority of methods is marginally inadequate.

Anomaly detection can also be viewed as a one-class classification (OCC) problem, which has inspired some researches. As depicted in Fig. 3, the method finds a hypersphere to distinguish normal sample features from abnormal sample features during training. During inference, the method determines whether the sample is abnormal based on the relative position of the test sample's features and the hypersphere. Since the training set does not contain abnormal samples, some methods create abnormal samples artificially to improve the accuracy of the hypersphere.

SVDD[34] is a classic algorithm in the OCC problem, PatchSVDD[19], DSPSVDD[20] and SE-SVDD[21] improve it for industrial image AD. PatchSVDD[19] divides the image into uniform patches and sends them to the model for training, which significantly enhances the model's ability to detect anomalies. DSPSVDD[20] designs an improved comprehensive optimization objective for the deep SVDD model that simultaneously considers hypersphere volume minimization and network reconstruction error minimization to extract deep data features more effectively. SE-SVDD proposes a semantic correlation module (SCB) to improve the representation of abnormal semantics and the accuracy of anomaly localization by extracting multi-level features.

MOCCA[22] employs multi-layer features for anomaly detection. MOCCA, unlike SE-SVDD, uses an autoencoder to extract features and locates the boundary position of normal features at each layer. Sauter et al.[23] attempt to use the Xception network for classification and obtain results comparable to SVDD. FCDD[35] employs a fully convolutional neural network for OCC. Since the relative positions of the features of each image layer do not change during the convolution process, FCDD yields more interpretable results than alternative methods.

PANDA[25] examines the migration method of pre-

Table 3 A summary of one-class classification methods regarding loss function, pre-trained model, and highlights

Method	Loss function	Pre-trained	Highlights
Patch SVDD[19]	Cross-entropy, SVDD	–	The paper divides image into patches and sends them to SVDD for training.
DDSPSVDD[20]	L_2, SVDD	VGG	DSPSVDD takes reconstruction error into model training.
SE-SVDD[21]	SVDD	ResNet	The paper proposes a semantic correlation block (SCB) to represent abnormal semantics information.
MOCCA[22]	L_2, SVDD	–	The paper extends a single boundary to a hard boundary and a soft boundary, it also trains AE as feature exactor.
Sauter et al.[23]	Cross-entropy	Xception[24]	The paper uses Xception to train a classification network.
PANDA[25]	SVDD, log-likelihood	DN2[26]	The paper introduces a method to avoid combating collapse in model adaption.
Sohn et al.[27]	Cross-entropy, contrastive	ResNet	The paper presents a novel distribution-augmented contrastive learning to enhance the representing ability of network.
Bai et al.[28]	–	–	The paper performs template matching on salient regions to detect anomalies.
Niu et al.[29]	L_1, L_2	–	This paper uses saliency detection to obtain object contours to assist anomaly detection.
UISDI[30]	L_1, L_2, log-likelihood	–	The paper uses salient object detection to segment the foreground and foreground to obtain abnormal regions.
CutPaste[31]	Cross-entropy	EfficientNet	The paper applies "cut and paste" augmentation into binary anomaly classification.
Yoa et al.[32]	Cosine similarity, contrastive	–	The paper applies some dynamic local augmentation to generate negative samples.
CPC-AD[33]	InfoNCE	–	The paper applies contrastive predictive coding (CPC) model to AD and gets an anomaly score through pixel-wise loss.
MemSeg[18]	L_1, focal	ResNet	The paper artificially creates anomalies in the foreground of products and makes detecting artificial anomalies a segmentation task.

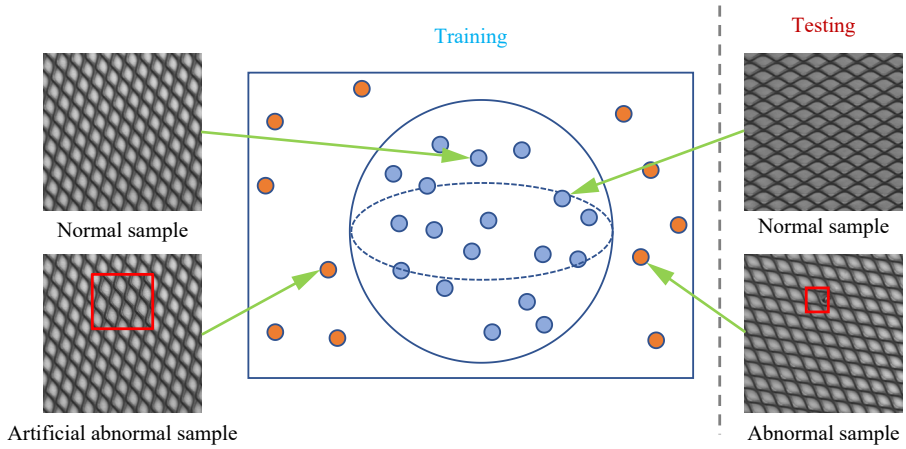

Fig. 3 Architecture of one-class classification models

trained features and introduces the early stopping mechanism to the OCC problem. In addition, Reiss and Hoshen[36] investigate the issue of catastrophic forgetting in PANDA. They propose a new loss function capable of overcoming the failure modes of both center-loss and contrastive-loss methods and replacing Euclidean distance with a confidence-invariant angular center loss for prediction.

DisAug CLR[27] proposes a two-stage anomaly detection framework, in which the first stage hinders the uniformity of contrastive representations by means of a novel distribution-enhanced contrastive learning. After comparative learning, abnormal and normal sample representations are easier to distinguish. While the second stage builds a one-class classifier using the representations learned in the first stage. Yoa et al.[32] present a novel dynamic local augmentation to generate negative image pairs from a normal training dataset, which is effective for anomaly detection. Contrastive predictive coding (CPC)[37] model is utilized by de Haan et al.[33] for anomaly detection and segmentation, which uses patch-wise contrastive loss as anomaly score to localize anomalies.

In addition, inspired by saliency object detection[38–40], many methods apply saliency detection to anomaly detec-

tion. Bai et al.[28] propose to use Fourier transform to detect salient regions of images, and compare the salient regions with templates to detect anomalies. Niu et al.[29] use the method of salient object detection to obtain object contours, thereby assisting the detection of outliers. Qiu et al.[30] propose a multi-scale saliency detection (MSSD) method to separate the foreground and background to obtain coarse anomaly regions, and refine the detected results on this basis. What's more, GradCAM[41], as a common method to obtain saliency maps, is also used in various anomaly detection algorithms. Both CutPaste[31] and CAVGA[42] treat anomaly detection as a classification problem, while GradCAM is used for pixel-level anomaly localization.

CutPaste[31] is a representative example of an OCC method for data augmentation. It generates abnormal images by cutting and pasting portions of normal images, allowing the network to distinguish abnormal images. Additionally, segmentation-based methods are useful. This method puts more emphasis on pixel-level anomaly localization. When the flow is known, Iquebal and Bukkapatnam[43] demonstrate that the maximum posterior estimation of image labels can be formulated as a continuous max-flow problem. Then, anomaly segmentation is accomplished by obtaining flows iteratively using a novel Markov random field on the image domain. The technique shows its adaptability using a dataset for metal additive manufacturing anomaly detection[44]. MemSeg[18] stores the features of normal images in a memory bank in order to improve the segmentation network's ability to distinguish abnormal regions. In order to prevent the influence of background factors, MemSeg only introduces anomalies in external datasets in the foreground of items, which is another reason for its excellent performance.

2.1.3 Distribution map

Distribution-map based methods necessitate a suitable mapping objective for training, and the choice of mapping method impacts model performance. As shown in Table 4, Normalizing flows (NF)-based methods predominate. As a generative model, NF has a strong mapping ability, and it has also demonstrated good performance in AD tasks.

Distribution-map based methods are very similar to OCC-based methods, with the exception that OCC-based methods concentrate on finding feature boundaries, whereas mapping-based methods attempt to map features into desired distributions. A common framework for those methods is shown in Fig. 4. This expected distribution is typically a multivariate Gaussian (MVG) distribution. This type of method first employs a strong pre-trained network to extract the features of normal images, and then maps the extracted features to the Gaussian distribution using a mapping module. This distribution will be deviated from the features of abnormal images that appear during the evaluation. The abnormal probability can be calculated based on the level of deviation.

Tailanian et al.[45] propose a contrario framework that applies statistical analysis to feature maps produced by patch PCA and ResNet, which performs well on leather samples, to detect anomalies in images. By fitting a mul-

Table 4 A summary of distribution-map based methods regarding loss function, pre-trained model, and highlights

Method	Loss function	Pre-trained	Highlights
Tailanian et al.[45]	PCA	ResNet	The paper uses PCA and ResNet to extract features and count their distribution.
Rippel et al.[46]	Cross-entropy	ResNet, efficientNet	The paper establishes a model of normality by fitting a multivariate Gaussian to feature representations of a pre-trained network.
Rippel et al.[47]	Mahalanobis distance	EfficientNet	The paper generates a multivariate Gaussian distribution for the normal class and mitigates the catastrophic forgetting in past research.
PEDENet[48]	Log-likelihood, cross-entropy, regularization	–	The model can predict the location of the patch and compare it with the actual location to judge the abnormality.
PFM[49]	L_2	ResNet	The paper proposes the bidirectional and multi-hierarchical bidirectional pre-trained feature mapping based on the vanilla feature mapping.
PEFM[50]	L_2	ResNet	The paper introduces position encoding into PFM.
FYD[51]	L_2	ResNet	The paper aligns samples at image and feature levels to detect anomalies.
DifferNet[52]	Log-likelihood	ResNet	The paper is the first one to introduce normalizing flow into anomaly detection.
CS-Flow[53]	Log-likelihood	ResNet	The paper uses information of multi-scale feature maps and improves DifferNet.
CFlow-AD[54]	Log-likelihood	ResNet	The paper introduces positional encoding into the conditional normalizing flow framework.
CAINNFlow[55]	Log-likelihood	ViT[56]	The paper uses VIT to replace ResNet and achieve better result.
FastFlow[57]	Log-likelihood	ResNet	The paper introduces an alternate stacking of large and small convolution kernels in the NF module to model global and local distribution.
AltUB[58]	Log-likelihood	ResNet	The paper designs a module for normalizing flow based methods and improves their performance.

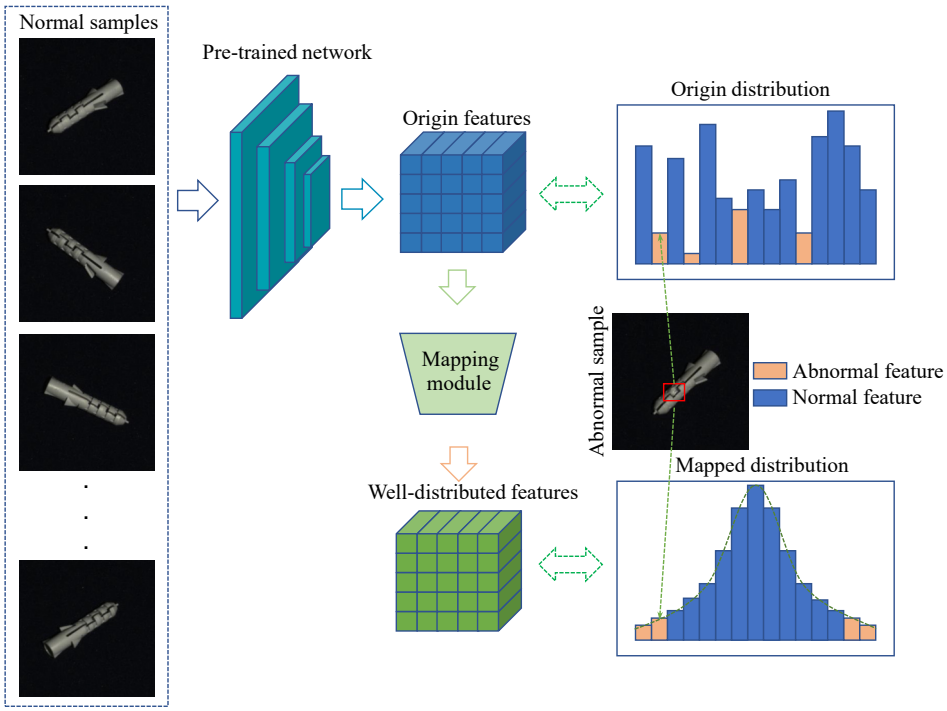

Fig. 4 Architecture of distribution-map based methods

tivariate Gaussian to the feature representations of a pre-trained network, Rippel et al.[46] establish a model of normality. Nonetheless, the issue of catastrophic forgetting remains unresolved. Based on the relationship between generative and discriminative modeling, Rippel et al.[47] generate a multivariable Gaussian distribution for the normal class and prove the efficacy of this concept on deep SVDD and FCDD, which mitigates the catastrophic forgetting observed in previous research. PEDENet[48] framework consists of a patch embedding (PE) network, a density estimation (DE) network, and a location prediction (LP) network. At first, the PE module is used to reduce the size of the features that the pre-trained network has extracted. Then, using the DE module, which was inspired by the Gaussian mixture model, and the LP module, the model can predict the relative position of the patch embedding and, based on the difference between the predicted result and the actual result during inference, decide if the image is abnormal. Pre-trained feature mapping (PFM)[49] proposes bidirectional and multi-hierarchical bidirectional pre-trained feature mapping to enhance the performance of vanilla feature mapping. In addition, Wan et al.[50] add position encoding to the PFM framework and propose a novel position encoding enhanced feature mapping (PEFM)[50] to further enhance PFM. FYD[51] introduces registration to industrial image AD for the first time. FYD suggests a coarse-to-fine alignment method that starts with aligning the foreground of objects at the image level. Next, in the refinement alignment stage, non-contrastive learning is used to increase the similarity of features between all corresponding positions in a batch.

Normalizing flows (NF)[59] is a technique for constructing complex distributions by transforming a probability density via a series of invertible mappings. NF methods extract features from normal images from a pre-trained model, such as ResNet[7] or Swin Transformer[60], and transform the feature distribution as a Gaussian distribution during the training phase. In the test phase, after passing through NF, the features of abnormal images will deviate from the Gaussian distribution of the training phase, which is the most important principle for classifying anomalies. DifferNet[52] is the first research to use NF to address the industrial image AD issue. By incorporating cross-convolution blocks within the normalizing flow to assign probabilities, CS-Flow[53] makes use of the context within and between multi-scale feature maps to improve DifferNet. CFlow-AD[54] adds positional encoding to the framework for conditional normalizing flow to achieve superior results. In addition, CFlow-AD[54] analyzes in depth why the multivariate Gaussian assumption is a reasonable priority in earlier models and why the more general NF framework aims to converge to similar results with less computation. FastFlow[57] introduces an alternate stacking of large and small convolution kernels in the NF module to model global and local distribution efficiently. CAINNFlow[55] enhances the performance of the model by introducing the attention mechanism CBAM[61] to the NF module. In techniques such as FastFlow and CFlow-AD, the feature distribution center is not 0 and their performance is unstable. Kim et al.[62] propose a simple solution AltUB[58] that uses alternating training to update the base distribution of normalizing flow for anomaly detection in order to solve the problem. AltUB

verifies the effect of CFlow-AD and FastFlow using AltUB.

2.1.4 Memory bank

As illustrated in Table 5, memory-based methods regularly do not require the loss function for training, and models are constructed quickly. Their performance is ensured by a robust pre-training network and additional memory space, and this type of method is currently the most effective in IAD tasks.

The primary distinction between memory bank-based methods and OCC-based methods, is that memory-based methods, such as SVDD, require additional memory space to store image features. As shown in Fig. 5, these methods require minimal network training and only require sampling or mapping the collected normal image features for inference. During inference, features of the test image are compared to features in the memory bank. The abnormal probability of the test image is equal to the spatial distance from the normal features in the memory bank.

K nearest neighbors (KNN)[72] is a widely used algorithm for unsupervised anomaly detection, but it operates only at the sample level. Semantic pyramid anomaly detection (SPADE)[63] is inspired by KNN and utilizes correspondences based on a multi-resolution feature pyramid to obtain pixel-level anomaly segmentation results. PaDim[73] employs multivariate Gaussian distributions to construct a probabilistic representation of the normal class. Consequently, the memory bank size is determined solely by the image resolution and not by the size of the training set. PaDiM requires the batch-inverse of the multidimensional covariance tensor, which makes it challenging to scale up to larger CNNs due to the increased feature size. To reduce the computational cost of the inverse by a factor of three, Kim et al.[62] generalize random feature selection into semi-orthogonal embedding.

Meanwhile, self-organizing map for anomaly detection (SOMAD)[64] and GCPF[65] enhance the storage of normal features. SOMAD preserves normal characteristics by employing topological memory based on multi-scale features. While GCPF transforms standard characteristics into multiple independent multivariate Gaussian clustering.

PatchCore[68] is a significant advancement in industrial image AD that significantly raises the performance for MVTec AD. Patchcore contains two special points. First, the memory bank of patchcore is coreset-subsampled to ensure a low inference cost while maximizing performance. Patchcore then determines whether the test sample is abnormal based on the distance between the test sample's nearest neighbor feature in its memory bank and other features. This process of reweighting makes patchcore more robust. Since patchcore was proposed, numerous improved methods have been developed on its foundation. Coupled-hypersphere-based feature adaptation (CFA) is proposed by Lee et al.[69] to obtain target-oriented features. The center and surface of the hypersphere in the memory bank are obtained through transfer learning, and the positional relationship between the test feature and the coupled-hypersphere can be used to determine whether it is abnormal or not. FAPM[70] is comprised of numerous patch-wise and layer-wise memory banks located in various places. FAPM calculates the features in different memory banks independently during inference, which significantly accelerates inference speed. N-pad[74] allows for the possibility of marginal misalignment

Table 5 A summary of memory bank based methods regarding loss function, pre-trained model, and highlights

Method	Loss function	Pre-trained	Highlights
SPADE[63]	–	ResNet	The paper uses multi-resolution feature to detect anomalies based on KNN.
Kim et al.[62]	–	ResNet	The paper reduces the computational cost for the inverse of multi-dimensional covariance tensor so that bigger resolution image can be applied.
SOMAD[64]	–	ResNet	The paper maintains normal characteristics by using topological memory based on multi-scale features.
GCPF[65]	–	ResNet	The paper processes normal features into multiple independent multivariate Gaussian clustering.
MSPB[66]	Kmeans, cosine similarity, SVDD	VGG	The paper enhances network representation capabilities by learning patch position relationships.
SPD[67]	Focal, InfoNCE, SPD, cosine similarity	–	Design a contrastive learning method to retrain ResNet to enhance the ability of defect representation.
PatchCore[68]	–	ResNet	The paper introduces a core-set sampling method to build a memory bank.
CFA[69]	SVDD	ResNet	The paper improves patchcore so that image features are distributed on a hypersphere.
FAPM[70]	–	ResNet	The paper puts different position features of the image into different memory banks to speed up retrieval.
N-pad[71]	Mahalanobis distance, log-likelihood	ResNet	The paper allows for possible edge misalignment by estimating a nominal distribution for each pixel using the pixel's neighborhood features.

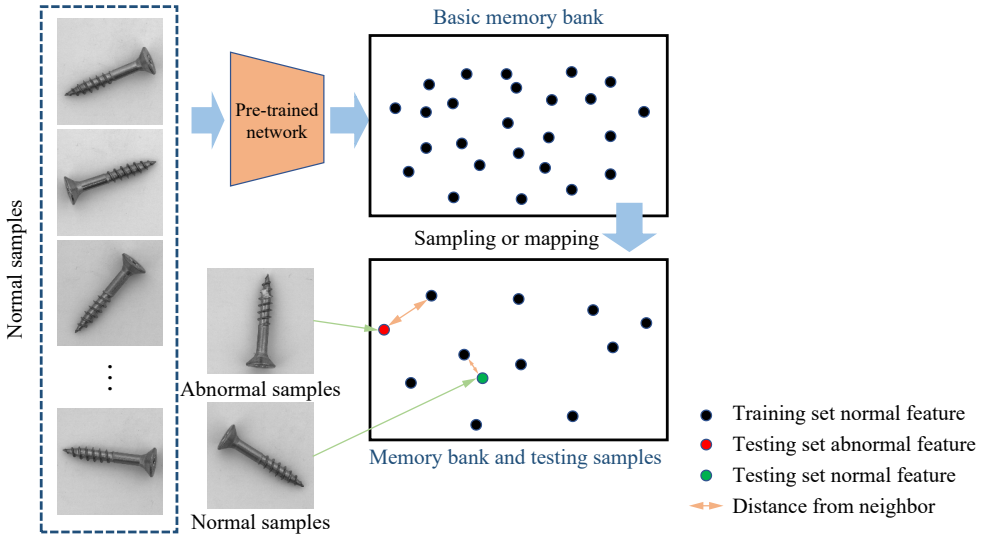

Fig. 5 Architecture of memory bank based methods

by estimating a per-pixel nominal distribution using neighboring and target pixel features. In addition, anomaly scores are deduced using both Mahalanobis and Euclidean distances between target pixels and the estimated distribution. Similarly, Bae et al.[71] model the cumulative histogram using location information as conditional probabilities, and neighborhood information was used to establish the normal feature distribution. Furthermore, this work introduces the first refinement approach in the anomaly detection and localization problem, using synthetic anomalous images to improve the anomaly map based on the input image, as well as using neighborhood and location information to estimate the distribution.

By learning the embedding position information and comparing the extracted features with the normal embedding during inference, Tsai et al.[66] propose a method to improve the network's ability to represent data. It is also based on the concept of self-supervised learning. Zou et al.[67] use contrastive learning to train the backbone network and propose a new data augmentation method called SPD to push the network to differentiate between two images with slight differences. In addition, they demonstrate the representation capability of the backbone network using patchcore[68].

2.2 Reconstruction based methods

Reconstruction-based methods primarily self-train encoders and decoders to reconstruct images for anomaly detection, which makes them less reliant on the pre-trained model and increases their ability to detect anomalies. However, its image classification capability is poor due to its inability to extract high-level semantic features. As shown in Table 6, the loss functions of various methods are comparable. However, their performance varies due to different reconstruction model paradigms and abnormal sample construction methods.

The structure of the reconstruction-based technique is depicted in Fig. 6. During the training process, normal or abnormal images are sent to the reconstruction network, and the reconstruction loss function is used to guide the training of the reconstruction network. Finally, the reconstruction network can restore the reconstruction image in a manner similar to the original normal image. In the inference stage, the comparison model compares the original image to the reconstructed image to generate a prediction. In contrast to the variety of methods for feature embedding, the majority of reconstruction-based methods only differ in the construction of the reconstruction network. Reconstruction-based methods outperform feature-embedding methods at the pixel level due to their ability to identify anomalies through pixel-level comparison. In addition, the majority of reconstruction-based methods are trained from scratch without employing robust pre-trained models, which results in inferior performance compared to image-level feature embedding.

2.2.1 Autoencoder

Autoencoder (AE) is the most prevalent reconstruction network for AD. Numerous other reconstruction networks also consist of encoder and decoder components. Bergmann et al.[75] investigate the influence of structure similarity index measure (SSIM) and L_2 loss on AE reconstruction and anomaly segmentation, providing numerous suggestions for future research.

How to resolve the difference between the reconstructed image and the original image is the most foundational principle. There are regularly differences in style between the reconstructed image and the original image, resulting in over-detection. Chung et al.[76] present an outlier-exposed style distillation network (OE-SDN) to preserve the style translation and suppress the content translation of the AE in order to avoid over-detection. As the anomaly prediction, Chung et al.[76] replace the difference between the original image and the reconstruction image of AE with the difference between the reconstruc-

Table 6 A summary of reconstruction based methods

Method	Loss function	Pre-trained	Highlights
(1) Autoencoder model			
Bergmann et al.[75]	L_2, SSIM	–	The paper firstly takes SSIM as a loss to reconstruct image and detect anomalies.
Chung et al.[76]	L_2, SSIM	–	The paper proposes two AEs and reduces style change during image reconstruction.
UTAD[77]	L_1, adversarial	VGG	The paper uses two-stage reconstruction to generate high-fidelity images to avoid reconstruction errors.
DFR[78]	L_2	VGG	The paper proposes to reconstruct and compare at the feature level to detect anomalies.
ALT[79]	L_1, perceptual, adversarial	VGG	The paper proposes an adaptive attention-level transition strategy and uses perceptual loss to improve reconstruction quality.
P-Net[80]	L_1, adversarial	–	The paper designs a new architecture for anomaly detection.
Collin and De Vleeschouwer[81]	L_1, L_2	–	The paper adds skip-connection in reconstruction network and adds noise during training to improve reconstruction sharpness.
Tao et al.[82]	L_2	VGG	The paper proposes a dense feature fusion module to assist reconstruction.
Hou et al.[83]	L_2, adversarial	–	The paper uses memory to help reconstructing images.
EdgRec[84]	L_2, SSIM	–	The paper reconstructs from the gray value edge and preserves the high-frequency information with skip-connection.
PAE[85]	L_2, cross-entropy	–	The paper gradually increases the resolution of the input image during training.
SMAI[86]	L_2, SSIM	–	The paper masks and inpaints image by superpixel.
RIAD[87]	L_2, MSGMS, SSIM	–	The paper proposes to inpaint and reconstruct images by patch.
I3AD[88]	L_1, adversarial	–	The paper gradually masks the high anomaly probability areas and reconstructs them.
Bauer [89]	L_2	–	The paper proposes to reconstruct the anomalous area differently from the original image.
Huang et al.[90]	L_2, SSIM, GMS	–	Similar to I3AD, but the paper adds skip connections to reconstruction network.
DREAM[91]	L_2, SSIM, focal	–	The paper designs a method to generate abnormal images and uses U-Net[92] to distinguish anomalies after reconstruction.
SGSF[93]	L_2, SSIM, focal	–	The method utilizes the idea of saliency detection to generate more realistic anomalies than DRAEM.
DSR[94]	L_2, focal	–	The paper generates abnormal samples in feature level and perform better than DRAEM.
NSA[95]	L_2, cross-entropy	–	The paper generates abnormal samples by pasting parts of other normal samples, which is the SOTA method without extra data.
SSPCAB[96]	L_2	–	The paper designs a "plug and play" self-supervised block to improve the reconstruction ability of many methods.
SSMCTB[97]	L_2	–	This paper replaces the SE-layer in SSPCAB with transformer architecture.
Dehaene et al.[98]	Cross-entropy	–	The paper guides reconstruction using gradient descent with VAE.
Liu et al.[99]	Attention disentanglement	–	The paper proposes to use disentanglement VAE to detect anomalies.
DGM[100]	L_2, log-likelihood	–	The paper proposes to use non-regularized objective functions for training VAE under heterogeneous datasets.
FAVAE[101]	Log-likelihood	VGG	The paper uses VAE to model the distribution of features extracted by its pre-trained model.
Wang et al.[102]	L_2, cross-entropy	–	The paper uses VQ-VAE to construct a discrete latent space and reconstructs images based on the latent space.
(2) GAN model			
SCADN[103]	L_2, adversarial	–	The paper masks part of image and reconstruct image with GAN during training.
AnoSeg[104]	L_1, L_2, adversarial	–	The paper generates abnormal samples through a GAN and detects anomalies with the discriminator.
OCR-GAN[105]	L_1, L_2, adversarial	–	The paper uses the frequency decoupling module to decouple and reconstruct images.
(3) Transformer model			
VT-ADL[106]	L_2, SSIM, log-likelihood	–	The paper proposes a transformer-based framework to reconstruct images and detects anomalies.

Table 6 (continued) A summary of reconstruction based methods

Method	Loss function	Pre-trained	Highlights
ADTR[107]	L_2, cross-entropy	EfficientNet	The paper makes it simple to identify anomalies when reconstruction fails by reconstructing features from pre-trained network.
AnoViT[108]	L_2	ViT	The paper uses a pre-trained ViT to extract features and reconstruct images.
HaloAE[109]	L_2, cross-entropy, SSIM	VGG	The paper introduces an auto-encoder architecture based on a transformer with HaloNet.
InTra[110]	L_2, GMS, SSIM	–	The paper leverages more global information to repair images with transformer.
MSTUnet[111]	L_2, SSIM, focal	–	The paper uses Swin Transformer for inpainting masked images and detects anomalies.
MeTAL[112]	L_1, SSIM	–	The paper uses information from neighbor patches to inpainting images, better accounting for local structural information.
UniAD[4]	L_2	EfficientNet	The paper trains all categories of products in one model.
(4) Diffusion model			
AnoDDPM[113]	L_2, log-likelihood	–	The paper is the first to apply diffusion model for industrial image anomaly detection.
[114]	L_2, log-likelihood	–	The paper significantly speeds up the inference process of anomaly detection using diffusion model.

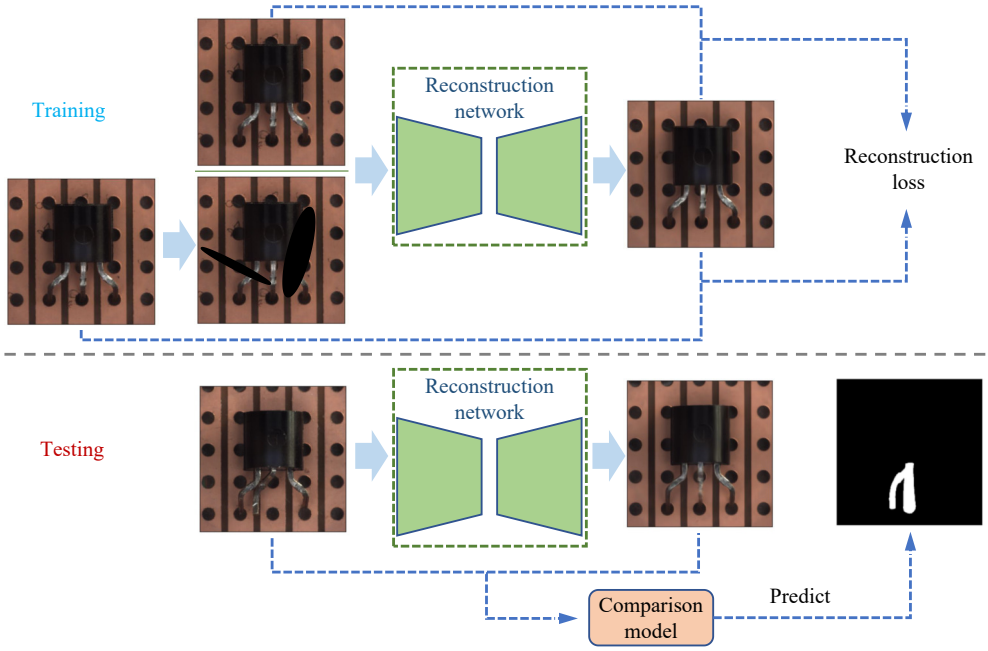

Fig. 6 Architecture of reconstruction based models

tion image of OE-SDN and the reconstruction image of AE. Unsupervised two-stage anomaly detection (UTAD)[77] brings an IE-Net and Expert-Net to extract and utilize impressions for anomaly-free and high-fidelity reconstructions, thereby offering the framework interpretable.

Reconstruction-based methods are nearly effective as feature embedding methods when utilizing features at different scales. Similar to teacher-student architecture, deep feature reconstruction (DFR)[78] method detects anomalous through reconstruction at the level of features. DFR obtains multiple spatial context-aware representations from a network that has been pre-trained. Then, DFR reconstructs features using a deep yet efficient convolutional AE and detects anomalous regions by comparing the original features to the reconstruction features. Yan et al.[79] propose a novel multi-level image reconstruction (MLIR) framework that forms the reconstruction process as an image denoising task at different resolutions. Thus, MLIR accounts for the detection of both global structure anomalies and detail anomalies.

Modifying the structure of AE can also improve its capacity for reconstruction. Zhou et al.[80] introduce P-Net to compare the difference in structure between the original and reconstruction images. Collin and De Vleeschouwer[81] include skip-connections between encoder and decoder to improve the reconstruction′s sharpness. In addition, they propose corrupting them with a synthetic noise

model to prevent the network from convergently mapping identities, and they introduce the innovative Stain noise model for this purpose. Tao et al.[82] also operate at the feature level. They employ a dense feature fusion module to obtain a dense feature representation of double input in order to help reconstruction in the dual-Siamese framework. Hou et al.[83] also use skip-connections to enhance the quality of reconstruction. In addition to achieving expected results, they add a memory module to skip-connections. Liu et al.[84] reconstruct the original RGB image from its gray value edges, with the skip-connections in the model preserving the image′s high-frequency information to better guide the reconstruction. Progressive autoencoder (PAE)[85] improves autoencoder reconstruction performance through progressive learning and modified CutPaste augmentation. During training, PAE achieves progressive learning by gradually increasing the input image′s resolution.

Masking and repainting is an effective method for self-supervised learning. The superpixel masking and inpainting (SMAI) technique was developed by Li et al.[86] SMAI divides the image into multiple blocks of superpixels and trains the inpainting module to reconstruct a superpixel within a mask. SMAI performs masking and inpainting superpixel-by-superpixel on the test image during inference, and then compares the reconstruction image to the test image to distinguish abnormal regions. Iterative image inpainting anomaly detection (I3AD) is a method proposed by Nakanishi et al.[88] that reconstructs partial regions based on the anomaly map. I3AD improves reconstruction quality by only reconstructing inpainting masks over images, and only masking regions with a high probability of abnormality. SSM[90] is conceptually similar to I3AD. SSM adds skip-connections to the reconstruction network and predicts the mask region as the training target. RIAD[87] randomly masks a portion of the training set image at the patch level and reconstructs it using a U-Net encoder-decoder network[92]. During inference, RIAD combines multiple random masks and reconstruction patches to generate a reconstructed image, which is then compared to the original image. Multi-scale gradient magnitude similarity (MSGMS) outperforms SSIM as an anomaly score, according to RIAD.

DRAEM[91] is representative of reconstruction-based techniques. DRAEM synthesizes abnormal images and reconstructs them as normal by introducing external datasets, which greatly improves the reconstruction network′s generalization capacity. In addition, DRAEM feeds the original image and the reconstructed image into the segmentation network to predict abnormal regions, significantly enhancing the model′s ability to segment anomalous regions. Nevertheless, DRAEM is susceptible to failure when synthesizing near-in-distribution anomalies. Inspired by saliency detection, Xing et al.[93] propose the saliency augmentation module (SAM) to generate more realistic abnormal images than DRAEM, so as to achieve better results. DSR[94] proposes an architecture based on quantized feature space representation and dual decoders to circumvent the requirement for image-level anomaly generation. By sampling the learned quantized feature space at the feature level, the near-in-distribution anomalies are generated in a controlled way. NSA[95] does not use external data for data augmentation and adopts more data augmentation methods, allowing it to outperform all previous methods that learned without utilizing additional datasets. In contrast to other methods that attempt to reconstruct abnormal images into normal images, Bauer[89] proposes reconstructing the abnormal areas of the image so that they deviate from the original image′s appearance. This approach produces comparable results to other methods.

In contrast to classical reconstruction-based methods, Ristea et al.[96] propose integrating reconstruction-based functionality into a self-supervised predictive convolutional attentive block (SSPCAB). SSPCAB can be incorporated into models such as DRAEM and CutPaste to enhance those models. Self-supervised masked convolutional transformer block (SSMCTB)[97] transforms the SE-layer[115] in SSPCAB into a channel-wise transformer block and achieves superior results.

VAE is a variant of AE, with the difference that the intermediate variables of VAE are data from a normal distribution. Naturally, VAE has superior interpretability. Dehaene et al.[98] iteratively guide reconstruction using gradient descent with energy defined by the reconstruction loss, thereby overcoming the tendency of VAE to produce blurry reconstructions and preserving the normal high-frequency structure. The variational autoencoder is trained with an attention disentanglement loss by Liu et al.[99] Anomaly inputs in this VAE will result in Gaussian-deviating latent variables during gradient backpropagation and attention generation. This deviation can be used to locate anomalies. According to Matsubara et al.,[100] datasets are commonly heterogeneous rather than regularized, and non-regularized objective functions are more suitable for training VAE models on heterogeneous datasets. FAVAE[101] employs VAE to model the distribution of features extracted by the pre-trained model, implicitly simulating richer anomalies and enhancing the model′s generalization. Wang et al.[102] use VQ-VAE to create a discrete latent space, resample the discrete latent code that deviate from the normal distribution, and reconstruct the image using the resampled latent code. VQ-VAE reconstructs images that are closer to the training set′s normal images.

2.2.2 Generative adversarial networks

The stability of the reconstruction model based on generative adversarial networks (GANs) is not as good as that of AE, but the discriminant network has a better effect on some scenes described as follows.

During training, semantic context based anomaly detection network (SCADN)[103] masks a portion of the im-

age and reconstructs it with GAN. SCADN detects anomalies for inference by comparing the input image to the reconstruction image. In addition to masking images, AnoSeg[104] utilizes hard augmentation, adversarial learning, and channel concatenation to generate abnormal samples. AnoSeg then trains GAN to generate normal samples. AnoSeg differs from the AE reconstruction model in that its objective function incorporates both reconstruction loss and adversarial loss. OCR-GAN[105] utilizes the frequency decoupling (FD) module to decouple the image into information combinations of different frequencies, and then reconstructs and combines the information of these different frequencies to yield reconstructed images. During inference, the model can identify a statistically significant difference between the frequency distributions of normal and abnormal images.

2.2.3 Transformer

Transformer has a higher capacity to represent global information, which gives it the potential to surpass AE and become a new reconstruction network foundation for anomaly detection. Mishra et al.[106] propose a transformer-based framework to reconstruct images at the patch level and employ a Gaussian mixture density network to localize anomalous regions. You et al.[107] propose ADTR for reconstructing pre-trained features. According to them, the use of transformers prevents well-reconstructed anomalies, making it easy to identify anomalies when reconstruction fails. Lee and Kang[108] introduce a vision transformer-based encoder-decoder model (AnoViT) and assert that AnoViT is superior to the CNN-based L_2-CAE in the issue of anomaly detection. HaloAE[109] implements transformer into HaloNet[116] and facilitates image reconstruction by reconstructing features to achieve competitive results on the MVTec AD dataset. A common self-supervised learning method for reconstruction-based anomaly detection is the reconstruction of masked images. However, traditional CNNs find it difficult to extract global context information. In order to accomplish this, Pirnay and Chai[110] propose inpainting transformer (InTra), which integrates information from larger regions of the input image. InTra is representative of trained-from-scratch methods. Masked Swin Transformer Unet (MSTUnet)[111] is comparable to InTra, but MSTUnet employs additional enhancements[117] when simulating anomalies, thereby achieving superior results. De Nardin et al.[112] use the neighbor patch to reconstruct the masked patch and also achieve a powerful reconstruction ability.

2.2.4 Diffusion model

Diffusion model[118] is a recently popular generative model that can also be utilized for reconstruction-based anomaly detection. AnoDDPM[113] is, to the best of our knowledge, the first to apply the diffusion model to industrial image anomaly detection. In comparison to GAN-based methods, AnoDDPM with simplex noise can also capture large anomaly regions without the need for large datasets. When applying the diffusion model to anomaly detection, Teng et al.[114] primarily make two improvements. As a replacement metric for reconstruction loss, a time-dependent gradient value of normal data distribution is used to measure the defects. In addition, they develop a novel T-scales method to reduce the required number of iterations and accelerate the inference process.

3 Supervised anomaly detection

Despite the fact that abnormal data is diverse and difficult to collect, it is still possible to collect abnormal samples in real-world scenarios. Therefore, some researches focus on how to train models for anomaly detection using a small number of abnormal samples and a large number of normal samples.

Chu and Kitani[119] propose a semi-supervised framework for detecting anomalies in the presence of significant data imbalance. They assume that changes in loss values during training can be used to identify abnormal data as features. To achieve this, they train a reinforcement learning-based neural batch sampler to amplify the difference in loss curves between anomalous and non-anomalous regions. FCDD[35] is an unsupervised method that synthesizes abnormal samples for training the OCC model. This concept is transferable to other OCC methods. Venkataramanan et al.[42] propose a convolutional adversarial variational autoencoder (CAVGA) with guided attention that can be applied equally to cases with and without abnormal images. In an unsupervised setting, CAVGA is guided to focus on all normal regions of an image by an attention expansion loss. CAVGA uses a complementary guided attention loss in the weakly supervised setting to minimize the attention map corresponding to abnormal regions of the image while focusing on normal regions. Božič et al.[120] examine the influence of image-level supervision information, mixed supervision information, and pixel-level supervision information on surface defect detection tasks within the same deep learning framework. Božič et al.[120] find that a small number of pixel-level annotations can help the model achieve performance comparable to full supervision. DevNet[121] uses a small number of abnormal samples to realize fine-grained end-to-end differentiable learning. Wan et al.[122] propose a logit inducing loss (LIS) for training with imbalanced data distribution and an abnormality capturing module (ACM) for characterizing anomalous features in order to effectively utilize a small amount of anomalous information. DRA[123] proposes a framework for learning disentangled representations of seen, pseudo, and latent residual anomalies in order to detect both visible and invisible anomalies.

Besides, a number of studies fail to account for the unbalanced distribution of normal and abnormal samples and rely primarily on abnormal samples for supervised training. Sindagi and Srivastava[124] investigate the do-

main transfer problem of datasets for anomaly detection in various settings. Dual weighted PCA (DWPCA) is an algorithm proposed by Qiu et al.[125] for image registration and surface defect detection. An interleaved deep artifacts-aware attention mechanism (iDAAM) is proposed by Bhattacharya et al.[126] that propose to classify multi-object and multi-class defects in abnormal images. Zeng et al.[127] view anomaly detection as a subset of target detection and design a reference-based defect detection network (RDDN) to detect anomalies using template reference and context reference. Song et al.[128] regard the abnormal part as the salient area of the image, and propose an effective saliency propagation algorithm for anomaly detection. Long et al.[129] investigate defect detection in a tactile image, which has obvious benefits for fabric structure defect detection in RGB images. In addition, there are methods that refer to the concept of semantic segmentation. To detect defects in infrared thermal volumetric data, Hu et al.[130] propose a hybrid multi-dimensional space and temporal segmentation model. Ferguson et al.[131] use mask region-based CNN architecture to detect and segment defects in X-ray images simultaneously. There are also numerous modified models on anomaly detection based on the object detection and semantic segmentation model of natural images under full supervision[132–134]. There are also many weakly supervised object detection methods suitable for anomaly detection[135–138]. Here we will not discuss them one by one.

4 Industrial manufacturing setting

This section introduces the classification standards or application settings that are more appropriate for industrial scenes, namely few-shot anomaly detection, noisy anomaly detection, anomaly synthesis, and 3D anomaly detection.

4.1 Few-shot anomaly detection

Few-shot learning is meaningful for data collection and data labeling, which has a great influence on real-world applications. On the one hand, by studying few-shot learning, we can reduce the cost of data collection and data annotation for industrial products. On the other hand, we can solve the problem from the perspective of data and investigate what kind of data is most valuable for industrial image anomaly detection. Few-shot anomaly detection (FSAD)[139, 140] is still in its infancy. There are two settings in FSAD. The first setting is meta-learning[141]. In other words, this setting requires a large amount of images as meta-training dataset. Wu et al.[139] propose a novel architecture, called MetaFormer, that employs meta-learned parameters to achieve high model adaptation capability and instance-aware attention to localize abnormal regions. RegAD[141] trains a model for detecting category-agnostic anomalies. In the test phase, the anomalies are identified by comparing the registered features of the test image and its corresponding normal images. The second setting relies on the vanilla few-shot image learning. PatchCore[68], SPADE[63] and PaDim[73] conduct the ablation study on 16 normal training samples. None of them, however, are specialized in few-shot anomaly detection. Hence, it is necessary to develop new algorithms that concentrate on native few-shot anomaly detection tasks.

Recently, researchers extended the zero-shot anomaly detection (ZSAD) setting beyond the FSAD setting. The goal of ZSAD is to leverage the generalization power of large models to solve anomaly detection problems without any training, thus completely eliminating the cost of data collection and annotation. MAEDAY[142] uses a pre-trained masked autoencoder (MAE)[143] to tackle the problem. MAEDAY randomly masks parts of an image and restores them using MAE. If the reconstructed region is different from the region before masking, this region is considered as anomalous. WinCLIP[144] utilizes another large model called CLIP[145] for ZSAD. Basically, WinCLIP uses the image encoder of CLIP to extract image features. Given the textual descriptions such as “a photo of a damaged object”, WinCLIP uses the text encoder of CLIP to extract the features of these descriptions, and then calculates the similarity between text features and image features. If the similarity is high, the image is “a photo of a damaged object”, otherwise the image is normal. MAEDAY and WinCLIP demonstrate that zero-shot anomaly detection is a promising research direction.

4.2 Noisy anomaly detection

Noisy learning is a classical problem for anomaly detection. By studying anomaly detection under noisy learning, we can avoid the performance loss caused by labeling errors and reduce false detection in anomaly detection. Tan et al.[146] employ a novel trust region memory update scheme to keep noise feature point away from the memory bank. Yoon et al.[147] use a data refinement approach to improve the robustness of one-class classification model. Qiu et al.[148] propose a strategy for training an anomaly detector in the presence of unlabeled anomalies, which is compatible with a broad class of models. They create labelled anomalies synthetically and jointly optimize the loss function with normal data and synthesis abnormal data. Chen et al.[149] introduce an interpolated Gaussian descriptor that learns a one-class Gaussian anomaly classifier trained with adversarially interpolated training samples. However, the majority of the aforementioned approaches have not been verified on real industrial image datasets. In other words, the effectiveness of the existing anomaly detection methods may not be suitable for industrial manufacturing.

4.3 3D anomaly detection

3D anomaly detection can utilize more spatial information, thereby detecting some information that cannot be contained in RGB images. In some special lighting environments or for some anomalies that are not sensitive to color information, 3D anomaly detection can demonstrate its significant advantages. This research direction is currently receiving significant attention in the academy. Since the release of MVTec 3D-AD[6] dataset, several papers have focused on anomaly detection in 3D industrial images. Bergmann and Sattlegger[150] introduce a teacher-student model for 3D anomaly detection. The teacher network is trained to acquire general local geometric descriptors by recreating local receptive fields. While the student network is taught to match the local 3D descriptors of the pre-trained teacher network. Horwitz and Hoshen[151] propose BTF, a method that combines hand-crafted 3D representations (FPFH[152]) with the representation method of 2D features (PatchCore[68]). Reiss et al.[153] propose that the representational ability of self-supervised learning is temporarily inferior to that of handcrafted features for 3D anomaly detection. Nevertheless, self-supervised characterization still has great potential if large-scale 3D anomaly detection datasets are available. AST[16] employs RGB image with depth information to enhance anomaly detection performance. However, most of 3D IAD methods are specialized in RGB-D images, while the 3D dataset in real-world industrial manufacturing consists of point clouds, meaning current 3D IAD methods cannot be directly deployed in industrial manufacturing. Thus, there are still opportunities for 3D IAD advancement.

4.4 Anomaly synthesis

By artificially synthesizing anomalies, we can improve the performance of models with limited data. This research is complementary to the few-shot research. Few-shot learning studies how to improve the model when the data is fixed, and this research studies how to artificially increase the credible data to improve the model performance when the model is fixed. Both of them can reduce the cost of data collection and labeling. There are many unsupervised anomaly detection works that use data augmentation to synthetic anomaly images and significantly improve model performance. For examples, CutPaste[31], DRAEM[91], MemSeg[18] are representative methods.

In addition, some supervised methods use limited abnormal samples to synthesize more abnormal samples for training. Liu et al.[154] propose a model designed to generate defects on defect-free fabric images for training semantic segmentation. While Rippel et al.[155] use CycleGAN[156] containing ResNet/U-Net as a generator as the basic architecture to transfer defects from one fabric to another. By improving the style transfer network, SDGAN[157] achieves better results than CycleGAN. Wei et al.[158] propose a model named DST to simulate defect samples. First, DST generates a blank mask area on a non-defective image, then DST uses the masked histogram matching module to make the color of the blank mask area consistent with the overall color of the image, and finally DST uses U-NET to perform style transfer to make the generated image more realistic. Wei et al.[159] propose a model named DSS, which uses conventional GAN to reconstruct defect structures in designated regions of defect-free samples, and then uses DST for style transfer to blend simulated defects into the background. Jain et al.[160] try to use DCGAN, ACGCN and InfoGAN to generate defect images by adding noise, which improves the accuracy of classification. Wang et al.[161] propose DTGAN based on StarGANv2, which adds front-background decoupling and achieves a certain degree of style control and uses the Frechet inception distance (FID[162]) and kernel inception distance (KID[163]) to evaluate the quality of image generation. DefectGAN[164] also believes that defects and normal backgrounds can be layered, and that defects are foreground. DefectGAN generates defect foregrounds and their spatial distribution in the form of style transfer. Although there is a considerable amount of research in this field, unlike other fields that have well-established directions, there is still significant potential for further development.

5 Datasets and metrics

Datasets. Data is a crucial driving factor for machine learning, particularly for deep learning. Principally, the difficulty of getting industrial photos hampers the advancement of image anomaly detection in industrial vision. Table 7 demonstrates that the number and the size of IAD dataset are gradually increasing, but most of them are not generated in a real production line. The promising alternative approach is to fully utilize the industrial simulator to generate anomalous images, possibly reducing the gap between academic research and the demands of industrial manufacturing.

Metrics. Table 8 offers a comprehensive review of the metrics in industrial image anomaly detection. The first column denotes the name of the metric and the second column denotes the level. In other words, if the level is up, the larger the metrics value, the better the performance. If the level is down, the lower the metrics value, the better the performance. The third column gives the detail for each metric, especially on how the metric accurately indicates the performance of image anomaly detection. From Table 8, it can be easily observed that most of novel metrics are the variants of natural image segmentation and detection metrics, such as F1 score, AU-ROC or AU-PR. However, these metrics can not correspond to the performance of IAD because the tiny size of anomalies requires a greater weighting than the anomaly-free

Table 7 Comparison of datasets for anomaly detection

Dataset	Class	Normal	Abnormal	Total	Annotation level	Real or synthetic
AITEX[165]	1	140	105	245	Segmentation mask	Real
BTAD[106]	3	–	–	2 830	Segmentation mask	Real
DAGM[166]	10	–	–	11 500	Segmentation mask	Synthetic
DEEPPCB[167]	1	–	–	1 500	Bounding box	Synthetic
Eycandies[168]	10	13 250	2 250	15 500	Segmentation mask	Synthetic
Fabric dataset[169]	1	25	25	50	Segmentation mask	Synthetic
GDXray[170]	1	0	19 407	19 407	Bounding box	Real
KolektorSDD	1	347	52	399	Segmentation mask	Real
KolektorSDD2[120]	1	2 979	356	3 335	Segmentation mask	Real
MIAD[171]	7	87 500	17 500	105 000	Segmentation mask	Synthetic
MPDD[172]	6	1 064	282	1 346	Segmentation mask	Real
MTD[173]	1	–	–	1 344	Segmentation mask	Real
MVTec AD[5]	15	4 096	1 258	5 354	Segmentation mask	Real
MVTec 3D-AD[174]	10	2 904	948	3 852	Segmentation mask	Real
MVTec LOCO-AD[6]	5	2 347	993	3 340	Segmentation mask	Real
NanoTwice[175]	1	5	40	45	Segmentation mask	Real
NEU surface defect database[176]	1	0	1 800	1 800	Bounding box	Real
RSDD[177]	2	–	–	195	Segmentation mask	Real
Steel defect detection[178]	1	–	–	18 076	Image	Real
Steel tube dataset[179]	1	0	3 408	3 408	Bounding box	Real
VisA[67]	12	9 621	1 200	10 821	Segmentation mask	Real

Table 8 A summary of metrics used for anomaly detection

Metric/level	Formula	Remarks/usage
Precision (P) ↑	$P = TP/(TP + FP)$	True positive (TP), false positive (FP)
Recall (R) ↑	$R = TP/(TP + FN)$	False negative (FN)
True positive rate (TPR) ↑	$TPR = TP/(TP + FN)$	
False positive rate (FPR) ↓	$FPR = FP/FP + TN)$	True negative (TN)
Area under the receiver operating characteristic curve (AU-ROC) ↑	$\int_0^1 (TPR)\,\mathrm{d}(FPR)$	Classification
Area under precision-recall (AU-PR) ↑	$\int_0^1 (P)\,\mathrm{d}(R)$	Localization, segmentation
Per-region overlap (PRO)[180] ↑	$PRO = \frac{1}{N}\sum_i \sum_k \frac{P_i \cap C_{i,k}}{C_{i,k}}$	Total ground truth number (N)/ Predicted abnormal pixels (P)/ Defect ground truth regions (C)/ Segmentation
Saturated per-region overlap (sPRO)[174] ↑	$sPRO(P) = \frac{1}{m}\sum_{i=1}^{m} \min(\frac{A_i \cap P}{s_i}, 1)$	Total ground truth number (m)/ Predicted abnormal pixels (P)/ Defect ground truth regions (A)/ Corresponding saturation thresholds (s) / Segmentation
F1 score ↑	$F1 = 2(P \times R)/(P + R)$	Classification
Intersection over union (IoU)[181] ↑	$IoU = (H \cap G)/(H \cup G)$	Prediction (H), ground truth (G)/ localization, segmentation

regions. Hence, the validity of these metrics for IAD remains to be explored.

6 Total performance analysis

Tables 9 and 10 show the statistical result of current IAD performance on MVTec AD. Fig. 7 supports the results of Table 9: Even if different methods have similar performance in image classification, there are still significant differences in pixel-level segmentation. We provide a deep analysis of the performance of current IAD methods and unlock meaningful insights as below:

1) Regarding the identification of image-level anomaly detection tasks, memory bank-based approaches are the most effective neural network design. However, they are inadequate at detecting pixel-level anomalies.

2) Ensemble learning can dramatically improve the performance of state-of-the-art anomaly detection methods.

3) SSPCAB[96] can be seamlessly integrated into cutting-edge methods and significantly enhance the performance of reconstruction-based methods.

4) The gap between few-shot IAD and vanilla IAD is narrowing. In other words, we may utilize data distillation algorithms to lower the amount of the dataset used for anomaly detection.

5) Without using ensemble learning, MemSeg[18] achieved the SOTA result on image-level anomaly classification, which is mainly due to the use of the U-Net[92] framework. DRAEM[91] also uses U-Net to outperform other methods on pixel-level anomaly segmentation. The effectiveness of MemSeg and DRAEM demonstrates the superiority of the segmentation module in anomaly detection. Artificial supervision is usually inferior to real supervision, and segmentation models trained with artificial supervision often perform worse. However, even when using artificially generated anomalies as supervisory information, these methods with segmentation modules still outperform other methods without segmentation modules on classification and segmentation tasks. We can conclude that the segmentation module is beneficial for anomaly detection tasks.

6) AU-PR is more valuable than AU-ROC for segmentation tasks[67]. As shown in Table 10, reconstruction-based methods outperform other methods on the pixel AU-PR metric. As for Fig. 7, the detection result of DRAEM is closest to the ground truth. It results in sharper edges and fewer false detection regions. We can infer from statistical data and visualizations that reconstruction-based methods are more suitable for segmentation tasks.

7 Future directions

We outline several intriguing future directions as follows:

1) We should build up a multi-modalities IAD dataset. In actual assembly lines, RGB images are insufficient to detect anomalies. Hence, we may employ additional modalities information, such as X-ray and ultrasound, to enhance anomaly detection performance.

2) Given that test samples are sequentially streamed on the product line, most IAD methods are incapable of making instantaneous predictions upon the arrival of a new test sample. In industrial manufacturing, the inference speed of IAD should be addressed in addition to its accuracy. Adopting multi-objective evolutionary neural architecture search algorithms to find the optimal trade-off architecture is thus a promising approach.

3) The majority of IAD methods use ImageNet pre-trained models to extract the features from industrial images, which inevitably results in the feature drift issue. Consequently, there is a pressing need to construct a pre-trained model for industrial images.

4) Most anomaly detection methods focus on the unsupervised setting. Although this setting can reduce the cost of data labeling, it greatly curbs the development of segmentation-based methods. Unsupervised methods and supervised methods should complement each other, and the main reason for the slow development of supervised methods in recent years is the lack of a large number of labeled data sets. Therefore, it is necessary to propose a fully supervised anomaly detection dataset with pixel-level annotations in the future.

5) Previously, we focused on developing data augmentation method for normal images. However, we have not made much effort on synthesizing abnormal samples via data augmentation. In industrial manufacturing, it is very difficult to collect a large number of abnormal samples since most of the production lines are faultless. Hence, more attention should be paid to abnormal synthesis methods in the future, like CutPaste[31], DRAEM[91] and MemSeg[18].

6) Current anomaly detection algorithms often focus on detection accuracy, while ignoring the storage size and efficiency of the models. This leads to high computation costs and limits the application of anomaly detection to the production end of enterprises. Therefore, it is necessary to design lightweight but efficient anomaly detection models.

7) Currently, image anomaly detection algorithms can be mainly categorized into two tasks: Industrial image anomaly detection and medical image anomaly detection. Although medical images have more modalities than industrial images[185–189], the two tasks share many similarities in terms of data and experimental settings. However, few studies have explored how to unify these two tasks. One reason for this is the domain differences between medical and industrial image datasets, and another reason is the lack of a good baseline and benchmark for comparison. It would be very meaningful to establish a unified framework for both industrial and medical image an-

Table 9 Image AUROC performance of different methods on MVTec AD. The highest and second places are marked in red and blue. All results are reported from the original papers.

Taxonomy	Method	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Avg.
Memory bank	PatchCore[68]	1.000	0.997	0.981	0.982	0.983	1.000	1.000	1.000	0.971	0.990	0.989	0.989	0.997	0.999	0.997	0.992
	PatchCore ensemble[68]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.996
	CFA[69]	1.000	0.998	0.973	0.973	0.992	1.000	1.000	1.000	0.979	0.973	0.994	1.000	1.000	0.997	0.996	0.993
	FAPM[70]	1.000	0.995	0.986	0.993	0.980	1.000	1.000	1.000	0.960	0.952	0.994	1.000	1.000	0.993	0.995	0.990
	N-pad[71]	1.000	0.995	0.994	0.993	0.987	1.000	1.000	1.000	0.980	0.974	1.000	1.000	0.996	0.996	0.993	0.994
	N-pad ensemble[71]	1.000	0.998	0.995	1.000	0.986	1.000	1.000	1.000	0.972	0.989	1.000	0.997	1.000	0.994	0.998	0.995
	MSPB[66]	1.000	0.988	0.972	0.934	1.000	0.996	0.993	0.978	0.977	0.941	0.962	1.000	0.989	0.997	0.995	0.981
	SPD[67]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.990
	SPADE[63]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.855
	[62]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.921
	SOMAD[64]	1.000	0.988	0.988	1.000	0.939	1.000	1.000	0.997	0.986	0.955	0.987	0.986	0.945	0.992	0.977	0.979
Teacher-student	RD4AD[14]	1.000	0.950	0.963	0.989	1.000	0.999	1.000	1.000	0.966	0.970	0.993	0.995	0.967	0.992	0.985	0.985
	STFPM[13]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.955
	Uninformed students[10]	0.918	0.865	0.916	0.695	0.819	0.937	0.819	0.895	0.935	0.928	0.912	0.863	0.701	0.725	0.933	0.857
	MKD[11]	0.994	0.892	0.805	0.793	0.780	0.984	0.951	0.736	0.827	0.833	0.916	0.922	0.856	0.943	0.932	0.877
	STPM[12]	1.000	0.996	0.930	0.987	1.000	0.998	1.000	1.000	0.981	0.968	0.999	0.979	0.983	0.993	0.993	0.987
	AST[16]	1.000	0.985	0.997	0.975	0.991	1.000	1.000	0.985	0.991	0.997	1.000	0.966	0.993	1.000	0.991	0.992
Distribution map	Rippel et al.[46]	0.998	0.955	0.938	1.000	0.817	0.996	0.997	0.947	0.884	0.854	0.998	0.964	0.963	0.986	0.978	0.953
	Rippel et al.[47]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.971
	PEDENet[48]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.928
	PFM[49]	1.000	0.988	–	1.000	0.980	1.000	1.000	1.000	0.965	0.918	0.996	0.886	0.978	0.995	0.974	0.975
	FYD[51]	1.000	0.953	0.925	0.988	0.989	0.999	1.000	0.999	0.945	0.901	0.988	1.000	0.992	0.994	0.975	0.977
	FastFlow[57]	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	0.994	0.978	1.000	0.944	0.998	1.000	0.995	0.994
	DifferNet[52]	0.990	0.959	0.869	0.929	0.840	0.993	0.971	0.961	0.888	0.963	0.994	0.986	0.911	0.998	0.951	0.949
	CS-Flow[53]	0.998	0.991	0.971	1.000	0.990	0.996	1.000	0.991	0.986	0.976	1.000	0.919	0.993	1.000	0.997	0.987
	CFLOW-AD[54]	0.989	0.975	0.988	0.990	0.988	0.990	0.996	0.988	0.984	0.991	0.965	0.988	0.952	0.950	0.991	0.982
	CS-Flow+AltUB[58]	1.000	0.978	0.981	0.992	1.000	1.000	1.000	0.995	0.970	0.917	0.999	0.994	0.952	0.990	0.985	0.984

Table 9 (continued) Image AUROC performance of different methods on MVTec AD. The highest and second places are marked in red and blue. All results are reported from the original papers.

Taxonomy	Method	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Avg.
One-class classification	Patch SVDD[19]	0.986	0.903	0.767	0.929	0.946	0.920	0.909	0.940	0.861	0.813	0.978	1.000	0.915	0.965	0.979	0.921
	SE-SVDD[21]	0.986	0.977	0.985	0.989	0.972	0.980	0.987	0.983	0.967	0.986	0.923	0.993	0.972	0.951	0.979	0.975
	MOCCA[22]	0.950	0.760	0.820	0.860	0.870	0.800	0.980	0.850	0.820	0.840	0.890	0.970	0.880	1.000	0.840	0.875
	PANDA[25]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.865
	Reiss and Hoshen[36]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.872
	Sohn et al.[27]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.865
	Yoa et al.[32]	0.918	0.883	0.965	0.894	0.881	0.962	0.985	0.926	0.964	0.972	0.919	0.958	0.883	0.892	0.954	0.930
	CPC-AD[33]	0.998	0.880	0.641	0.809	0.983	0.996	0.990	0.845	0.921	0.897	0.957	0.878	0.925	0.803	0.993	0.901
	CutPaste[31]	0.982	0.812	0.982	0.939	1.000	0.983	1.000	0.999	0.949	0.887	0.946	0.994	0.961	0.991	0.999	0.961
	MemSeg[18]	1.000	0.982	1.000	0.996	1.000	1.000	1.000	1.000	0.990	0.978	1.000	1.000	0.992	0.996	1.000	0.994
Reconst–AE	UTAD[77]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.900
	ALT[79]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.910
	Collin and De Vleeschouwer[81]	0.980	0.890	0.740	0.890	0.970	0.940	0.890	0.730	0.840	0.740	0.990	1.000	0.910	0.950	0.940	0.890
	Tao et al.[82]	1.000	0.983	0.916	0.968	0.956	0.994	0.918	0.977	0.895	0.981	0.964	1.000	0.913	0.983	0.961	0.961
	Hou et al.[83]	0.976	0.844	0.767	0.866	0.957	0.921	0.862	0.758	0.900	0.987	0.882	0.992	0.876	0.982	0.859	0.895
	EdgRec[84]	1.000	0.979	0.955	0.974	0.997	0.984	1.000	0.973	0.990	0.899	1.000	1.000	0.998	0.940	0.983	0.978
	PAE[85]	0.999	0.948	0.956	0.989	1.000	0.981	0.973	0.965	0.975	0.956	0.985	1.000	0.990	0.987	0.991	0.980
	SMAI[86]	0.860	0.920	0.930	0.880	0.970	0.970	0.860	0.920	0.920	0.960	0.620	0.960	0.850	0.800	0.900	0.890
	I3AD[88]	0.966	0.767	0.708	0.602	0.998	0.930	0.823	0.658	0.783	0.980	0.978	0.958	0.864	0.938	0.994	0.863
	Huang et al.[90]	0.999	0.773	0.914	0.763	1.000	0.915	0.999	0.887	0.891	0.850	0.944	1.000	0.910	0.959	0.999	0.920
	RIAD[87]	0.999	0.819	0.884	0.842	0.996	0.833	1.000	0.885	0.838	0.845	0.987	1.000	0.909	0.930	0.981	0.917
	DREAM[91]	0.992	0.918	0.985	0.970	0.999	1.000	1.000	0.987	0.989	0.939	0.996	1.000	0.931	0.991	1.000	0.980
	DSR[94]	1.000	0.938	0.981	1.000	1.000	0.956	1.000	0.985	0.975	0.962	1.000	0.997	0.978	0.963	1.000	0.982
	NSA[95]	0.977	0.945	0.952	0.956	0.999	0.947	0.999	0.987	0.992	0.902	1.000	1.000	0.951	0.975	0.998	0.972
	Bauer[89]	0.950	0.960	0.980	0.990	0.990	0.980	0.990	0.950	0.980	0.990	0.970	0.980	0.970	0.970	0.990	0.980
	DREAM+ SSPCAB[96]	0.984	0.969	0.993	0.982	1.000	1.000	1.000	1.000	0.998	0.979	1.000	1.000	0.929	0.995	1.000	0.989
	DREAM+ SSMCTB[97]	0.994	0.941	0.971	0.968	1.000	1.000	1.000	1.000	0.988	0.990	1.000	1.000	0.960	1.000	1.000	0.987

Table 9 (continued) Image AUROC performance of different methods on MVTec AD. The highest and second places are marked in red and blue. All results are reported from the original papers.

Taxonomy	Method	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Avg.
	NSA+SSPCAB[96]	0.977	0.956	0.954	0.975	0.999	0.942	0.999	0.990	0.992	0.911	1.000	1.000	0.956	0.977	0.998	0.975
	NSA+SSMCTB[97]	0.977	0.961	0.955	0.961	1.000	0.971	1.000	0.995	0.995	0.904	1.000	1.000	0.962	0.978	0.999	0.977
	FAVAE[101]	0.999	0.950	0.804	0.671	0.970	0.993	0.675	0.852	0.821	0.837	0.805	0.958	0.932	0.948	0.972	0.879
	Wang et al.[102]	0.990	0.720	0.680	0.710	0.910	0.940	0.960	0.830	0.680	0.800	0.950	0.920	0.730	0.960	0.970	0.850
Reconst–GAN	SCADN[103]	0.957	0.856	0.765	0.504	0.983	0.833	0.659	0.624	0.814	0.831	0.792	0.981	0.863	0.968	0.846	0.818
	Anoseg[104]	0.980	0.980	0.840	0.960	0.990	0.980	0.990	0.950	0.870	0.970	0.980	0.990	0.960	0.990	0.990	0.960
	OCR-GAN[105]	0.996	0.991	0.962	0.994	0.996	0.985	0.971	0.995	0.983	1.000	0.955	0.987	0.983	0.957	0.990	0.983
Reconst–transformer	ADTR[107]	1.000	0.925	0.925	1.000	0.978	0.999	1.000	0.945	0.933	0.942	1.000	0.939	0.980	0.999	0.970	0.969
	AnoViT[108]	0.830	0.740	0.730	0.500	0.520	0.880	0.850	0.860	0.720	1.000	0.890	0.740	0.830	0.950	0.730	0.780
	HaloAE[109]	1.000	0.846	0.884	0.697	0.951	0.998	0.978	0.884	0.901	0.896	0.957	0.972	0.844	1.000	0.997	0.914
	InTra[110]	1.000	0.703	0.865	0.988	1.000	0.957	1.000	0.969	0.902	0.957	0.982	1.000	0.958	0.975	0.994	0.950
	MSTUnet[111]	1.000	0.914	0.984	0.999	1.000	1.000	1.000	1.000	0.974	1.000	1.000	1.000	0.963	1.000	1.000	0.989
	MeTAL[112]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.863
Reconst–diffusion	VDD[114]	1.000	0.968	0.961	0.969	1.000	0.999	0.996	0.972	0.953	0.996	0.986	0.998	0.954	0.988	0.998	0.982
Few-shot	Metaformer[139]	0.991	0.971	0.875	0.940	0.859	0.994	–	0.962	0.901	0.975	0.990	1.000	0.944	0.992	0.986	0.958
	RegAD[141]	0.998	0.806	0.763	0.985	0.915	0.965	1.000	0.983	0.806	0.634	0.974	0.985	0.934	0.994	0.940	0.912
	Sheynin et al.[182]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.780
	RFS[140]	1.000	0.920	0.894	0.984	0.896	0.999	1.000	0.999	0.945	0.700	0.969	0.992	0.919	0.981	0.987	0.945
	MAEDAY[142]	0.937	0.690	0.649	0.979	0.839	0.941	0.999	0.734	0.817	0.614	0.984	0.925	0.753	0.995	0.943	0.853
Noisy	TrustMAE[146]	0.970	0.851	0.788	0.974	0.991	0.985	0.951	0.761	0.833	0.824	0.973	0.969	0.875	0.998	0.875	0.908
	Yoon et al.[147]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.937
	Qiu et al.[148]	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.959
	IGD[149]	1.000	0.997	0.915	0.913	0.958	0.873	0.946	0.828	0.991	0.978	0.906	0.906	0.997	0.825	0.970	0.934
	SoftPatch[183]	0.937	0.995	0.963	0.991	0.968	1.000	1.000	0.999	0.963	0.960	0.993	0.997	0.990	0.987	0.978	0.986
Supervised	CGVGA[42]	0.960	0.970	0.930	0.820	0.810	0.920	0.840	0.880	0.970	0.790	0.860	0.990	0.890	0.890	0.960	0.900
	DevNet[121]	0.993	0.892	0.865	0.867	0.967	1.000	0.999	0.991	0.866	0.970	0.987	0.860	0.924	0.999	0.990	0.945
	DRA[123]	1.000	0.909	0.935	0.940	0.987	1.000	1.000	0.997	0.904	0.977	0.994	0.826	0.915	0.998	1.000	0.959

Table 10 Pixel AUROC and AUPR performance of different methods on MVTec AD. The highest and second places are marked in red and blue. Note that * refers to reproduced results by us, while other results are reported from original papers.

Pixel AU-ROC																	
Taxonomy	Methods	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal_nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Average
Memory bank	PatchCore[68]	0.986	0.987	0.991	0.987	0.988	0.988	0.993	0.990	0.986	0.995	0.963	0.989	0.971	0.952	0.990	0.984
	FAPM[70]	0.982	0.985	0.990	0.989	0.978	0.986	0.990	0.982	0.980	0.990	0.980	0.987	0.982	0.940	0.986	0.980
	N-pad[71]	0.989	0.989	0.990	0.990	0.981	0.990	0.994	0.992	0.990	0.988	0.976	0.990	0.986	0.975	0.992	0.988
	Bae et al.[71]	0.990	0.991	0.993	0.994	0.989	0.992	0.995	0.993	0.990	0.994	0.983	0.991	0.984	0.965	0.993	0.989
	MSPB[66]	0.986	0.982	0.979	0.984	0.985	0.978	0.991	0.991	0.988	0.985	0.944	0.990	0.977	0.975	0.986	0.981
	SPADE[63]	0.984	0.972	0.990	0.975	0.937	0.991	0.976	0.981	0.965	0.989	0.874	0.979	0.941	0.885	0.965	0.960
	SOMAD[64]	0.983	0.982	0.987	0.989	0.984	0.984	0.991	0.980	0.980	0.991	0.948	0.985	0.953	0.944	0.987	0.978
	GCPF[65]	0.975	0.957	0.977	0.990	0.978	0.981	0.993	0.959	0.970	0.975	0.961	0.973	0.907	0.951	0.982	0.969
Teacher-student	MKD[11]	0.963	0.824	0.959	0.956	0.918	0.946	0.981	0.864	0.896	0.960	0.828	0.961	0.765	0.848	0.939	0.907
	Yamada and Hotta[13]	0.989	0.976	0.989	0.990	0.993	0.991	0.990	0.986	0.971	0.994	0.968	0.990	0.881	0.964	0.985	0.977
	Bergmann et al.[10]	0.993	0.983	0.985	0.992	0.996	0.995	0.996	0.989	0.987	0.993	0.988	0.993	0.907	0.981	0.992	0.985
	RD4AD[14]	0.987	0.974	0.987	0.989	0.993	0.989	0.994	0.973	0.982	0.996	0.956	0.991	0.925	0.953	0.982	0.978
	IKD[15]	0.990	0.980	0.986	0.987	0.970	0.987	0.985	0.984	0.988	0.986	0.957	0.986	0.971	0.939	0.976	0.978
Distribution based	PEDENet[48]	0.984	0.971	0.943	0.922	0.959	0.970	0.976	0.973	0.960	0.972	0.926	0.979	0.982	0.900	0.962	0.959
	PFM[49]	0.984	0.967	0.983	0.992	0.988	0.991	0.994	0.972	0.972	0.987	0.962	0.986	0.878	0.956	0.982	0.973
	PEFM[50]	0.985	0.983	0.985	0.992	0.992	0.992	0.994	0.970	0.970	0.990	0.966	0.992	0.984	0.965	0.986	0.983
	FYD[51]	0.983	0.975	0.986	0.985	0.968	0.987	0.992	0.982	0.973	0.987	0.968	0.989	0.981	0.996	0.982	0.982
	FastFlow[57]	0.977	0.984	0.991	0.994	0.983	0.991	0.995	0.985	0.992	0.994	0.963	0.989	0.973	0.970	0.987	0.985
	CFLOW-AD[54]	0.990	0.976	0.990	0.993	0.990	0.989	0.997	0.986	0.990	0.989	0.980	0.989	0.980	0.967	0.991	0.986
	CAINNFlow[55]	0.985	0.987	0.989	0.994	0.989	0.993	0.997	0.991	0.985	0.997	0.975	0.996	0.976	0.955	0.987	0.986
	CS-Flow+ AluUB[58]	0.990	0.976	0.990	0.993	0.991	0.989	0.997	0.986	0.990	0.989	0.980	0.989	0.982	0.966	0.991	0.987
	FastFlow+ AltUB[58]	0.990	0.984	0.991	0.995	0.993	0.993	0.997	0.987	0.991	0.995	0.976	0.992	0.980	0.969	0.991	0.988
One-class classification	Patch SVDD[19]	0.981	0.968	0.958	0.926	0.962	0.975	0.974	0.980	0.951	0.957	0.914	0.981	0.970	0.908	0.951	0.957
	SE-SVDD[21]	0.986	0.977	0.985	0.989	0.972	0.980	0.987	0.983	0.967	0.986	0.923	0.993	0.972	0.951	0.979	0.975
	CPC-AD[33]	0.890	0.840	0.720	0.740	0.800	0.810	0.940	0.760	0.770	0.650	0.820	0.810	0.900	0.820	0.950	0.820
	CutPaste[31]	0.976	0.900	0.974	0.983	0.975	0.973	0.995	0.931	0.957	0.967	0.905	0.981	0.930	0.955	0.993	0.960
	MemSeg[18]	0.993	0.974	0.993	0.992	0.993	0.988	0.997	0.993	0.995	0.980	0.995	0.994	0.973	0.980	0.988	0.988

Table 10 (continued) Pixel AUROC and AUPR performance of different methods on MVTec AD. The highest and second places are marked in red and blue. Note that * refers to reproduced results by us, while other results are reported from original papers.

Pixel AU-ROC																	
Taxonomy	Methods	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal_nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Average
Recons–AE	Chung et al.[76]	0.940	0.840	0.920	0.950	0.970	0.970	0.850	0.890	0.940	0.970	0.810	0.980	0.880	0.790	0.870	0.920
	DFR[78]	0.970	0.920	0.990	0.970	0.980	0.990	0.980	0.930	0.970	0.990	0.870	0.990	0.800	0.930	0.960	0.950
	ALT[79]	0.964	0.908	0.988	0.971	0.995	0.991	0.989	0.976	0.985	0.993	0.955	0.977	0.914	0.962	0.975	0.969
	AESc-Stain[81]	0.880	0.840	0.930	0.910	0.950	0.890	0.870	0.620	0.850	0.950	0.790	0.930	0.780	0.840	0.900	0.860
	Tao et al.[82]	0.964	0.971	0.983	0.991	0.981	0.988	0.992	0.983	0.967	0.993	0.909	0.986	0.870	0.941	0.982	0.967
	EdgRec[84]	0.983	0.977	0.952	0.994	0.992	0.994	0.997	0.980	0.987	0.977	0.986	0.992	0.943	0.914	0.987	0.977
	Kim et al.[85]	0.974	0.975	0.961	0.993	0.993	0.985	0.988	0.962	0.967	0.997	0.983	0.977	0.981	0.931	0.986	0.977
	I3AD[88]	0.950	0.795	0.854	0.850	0.987	0.756	0.938	0.526	0.725	0.959	0.788	0.969	0.651	0.776	0.962	0.832
	Huang et al.[90]	0.959	0.821	0.984	0.944	0.990	0.974	0.996	0.896	0.978	0.989	0.902	0.989	0.801	0.869	0.990	0.939
	RIAD[87]	0.984	0.842	0.928	0.963	0.988	0.961	0.994	0.925	0.957	0.988	0.891	0.989	0.877	0.858	0.978	0.942
	DRAEM[91]	0.991	0.947	0.943	0.955	0.997	0.997	0.986	0.995	0.976	0.976	0.992	0.981	0.909	0.964	0.988	0.973
	NSA[95]	0.983	0.960	0.976	0.955	0.992	0.976	0.995	0.984	0.985	0.965	0.993	0.949	0.880	0.907	0.942	0.968
	DRAEM+ SSPCAB[96]	0.988	0.960	0.931	0.950	0.995	0.998	0.995	0.989	0.975	0.998	0.993	0.981	0.870	0.968	0.990	0.972
	DRAEM+ SSMCTB[97]	0.992	0.955	0.934	0.958	0.997	0.995	0.976	0.993	0.974	0.995	0.993	0.990	0.891	0.948	0.990	0.972
	NSA+SSPCAB[96]	0.983	0.966	0.972	0.975	0.992	0.979	0.995	0.979	0.988	0.962	0.992	0.953	0.871	0.904	0.945	0.964
	NSA+SSMCTB[97]	0.984	0.975	0.979	0.956	0.992	0.979	0.996	0.983	0.984	0.964	0.991	0.954	0.883	0.935	0.947	0.967
	Dehaene et al.[98]	0.922	0.910	0.917	0.735	0.961	0.976	0.925	0.907	0.930	0.945	0.654	0.985	0.919	0.838	0.869	0.893
	Liu et al.[99]	0.870	0.900	0.740	0.780	0.730	0.980	0.950	0.940	0.830	0.970	0.800	0.940	0.930	0.770	0.780	0.860
	FAVAE[101]	0.963	0.969	0.976	0.960	0.993	0.987	0.981	0.966	0.953	0.993	0.714	0.987	0.984	0.899	0.968	0.953
	Wang et al.[102]	0.950	0.950	0.930	0.940	0.990	0.950	0.990	0.910	0.950	0.960	0.880	0.970	0.910	0.870	0.980	0.940
Recons–GAN	SCADN[103]	0.696	0.814	0.687	0.649	0.796	0.884	0.763	0.754	0.747	0.876	0.677	0.901	0.689	0.672	0.670	0.752
	Anoseg[104]	0.990	0.990	0.900	0.990	0.990	0.990	0.980	0.990	0.940	0.910	0.980	0.960	0.960	0.980	0.980	0.970
Recons–transformer	ADTR[107]	0.980	0.970	0.991	0.988	0.942	0.988	0.986	0.968	0.987	0.993	0.959	0.992	0.978	0.930	0.976	0.975
	AnoViT[108]	0.860	0.890	0.910	0.650	0.830	0.940	0.890	0.880	0.860	0.920	0.570	0.900	0.800	0.850	0.760	0.830
	HaloAE[109]	0.919	0.876	0.978	0.894	0.831	0.978	0.985	0.852	0.915	0.990	0.785	0.929	0.875	0.911	0.960	0.912
	InTra[110]	0.971	0.910	0.977	0.992	0.988	0.983	0.995	0.933	0.983	0.995	0.944	0.989	0.961	0.887	0.992	0.966
	MSTUnet[111]	0.990	0.899	0.957	0.983	0.997	0.993	0.995	0.993	0.976	0.974	0.997	0.991	0.746	0.980	0.989	0.964
	De Nardin et al.[112]	0.850	0.701	0.891	0.712	0.884	0.959	0.976	0.773	0.852	0.901	0.771	0.975	0.860	0.836	0.750	0.846

Table 10 (continued) Pixel AUROC and AUPR performance of different methods on MVTec AD. The highest and second places are marked in red and blue. Note that * refers to reproduced results by us, while other results are reported from original papers.

Pixel AU-ROC																	
Taxonomy	Methods	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal_nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Average
Recons–diffusion model	VDD[114]	0.979	0.975	0.986	0.989	0.997	0.992	0.993	0.979	0.960	0.996	0.944	0.983	0.952	0.951	0.993	0.978
Few-shot	Metaformer[139]	0.888	0.937	0.879	0.878	0.865	0.886	0.959	0.869	0.930	0.954	0.881	0.877	0.926	0.848	0.936	0.901
	MAEDAY[142]	0.959	0.842	0.953	0.982	0.966	0.983	0.994	0.684	0.913	0.974	0.901	0.922	0.860	0.929	0.962	0.922
Noisy	TrustMAE[146]	0.934	0.929	0.874	0.985	0.975	0.985	0.981	0.918	0.899	0.976	0.825	0.981	0.927	0.926	0.978	0.939
	SoftPatch[183]	0.975	0.971	0.989	0.989	0.974	0.924	0.993	0.983	0.976	0.969	0.954	0.985	0.936	0.929	0.986	0.969
Supervised AD	FCDD[35]	0.960	0.930	0.950	0.990	0.950	0.970	0.990	0.980	0.970	0.930	0.980	0.950	0.900	0.940	0.980	0.960
	DevNet[121]	0.951	0.920	0.938	0.963	0.935	0.959	0.990	0.876	0.859	0.897	0.950	0.819	0.839	0.900	0.973	0.918
Pixel AU-PR																	
Taxonomy	Methods	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal_nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Average
Memory bank	CFA*[184]	0.726	0.691	0.394	0.484	0.721	0.406	0.580	0.440	0.525	0.317	0.584	0.405	0.802	0.764	0.237	0.538
	PatchCore*[184]	0.768	0.653	0.442	0.627	0.325	0.537	0.456	0.870	0.777	0.354	0.546	0.372	0.610	0.477	0.595	0.561
	SPADE*[184]	0.699	0.234	0.265	0.573	0.321	0.418	0.531	0.438	0.518	0.236	0.662	0.482	0.793	0.452	0.437	0.471
	PaDiM$_*$[184]	0.730	0.343	0.334	0.497	0.580	0.374	0.452	0.394	0.602	0.249	0.517	0.406	0.713	0.423	0.166	0.452
Distribution based	FastFlow*[184]	0.688	0.276	0.418	0.205	0.539	0.455	0.351	0.276	0.633	0.311	0.259	0.525	0.609	0.378	0.045	0.398
Teacher student	RD4AD*[184]	0.787	0.527	0.451	0.574	0.492	0.621	0.482	0.791	0.784	0.536	0.532	0.518	0.549	0.482	0.571	0.580
	STPM*[184]	0.732	0.327	0.706	0.470	0.299	0.400	0.546	0.363	0.669	0.452	0.596	0.491	0.701	0.801	0.220	0.518
Recons–AE	EdgRec[84]	0.779	0.708	0.405	0.798	0.466	0.817	0.664	0.799	0.793	0.429	0.845	0.592	0.712	0.548	0.525	0.658
	DRAEM[91]	0.865	0.524	0.494	0.535	0.657	0.929	0.753	0.963	0.485	0.582	0.923	0.447	0.507	0.777	0.815	0.684
	DSR[94]	0.915	0.704	0.533	0.782	0.680	0.873	0.625	0.675	0.657	0.525	0.939	0.742	0.411	0.684	0.785	0.702
	DRAEM+ SSPCAB[96]	0.879	0.572	0.502	0.594	0.611	0.926	0.760	0.981	0.524	0.720	0.950	0.510	0.480	0.771	0.771	0.703
	DRAEM+ SSMCTB[97]	0.899	0.616	0.520	0.552	0.697	0.891	0.655	0.947	0.469	0.701	0.957	0.690	0.458	0.756	0.765	0.705

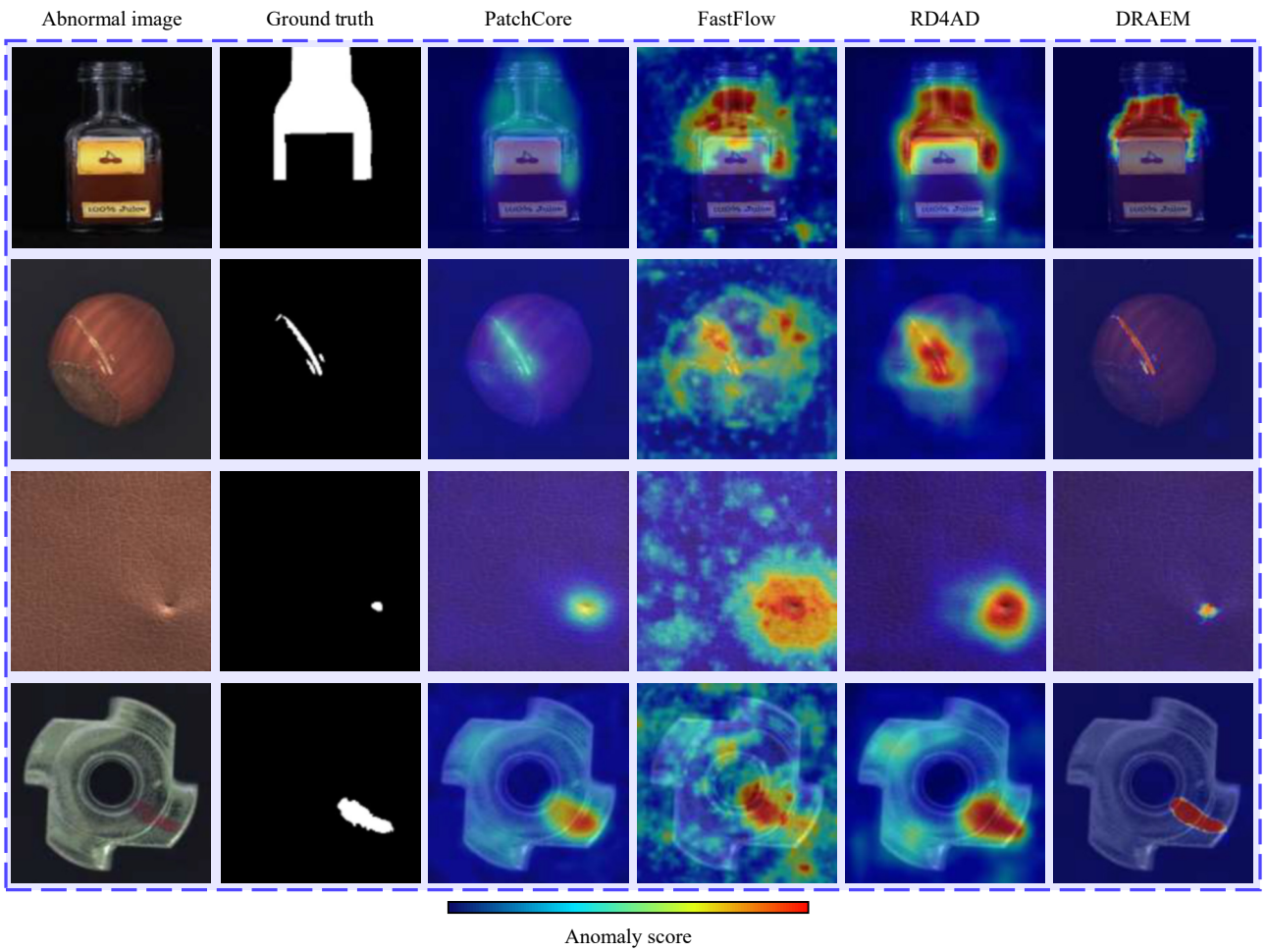

Fig. 7 Visualization of results from representative methods. Note that the visualization results are from the open-source code reproduction.

omaly detection at the data or method level.

8 Conclusions

In this paper, we provide a literature review on image anomaly detection in industrial manufacturing, focusing on the level of supervision, the design of neural network architecture, the types and properties of datasets and the evaluation metrics. In particular, we characterize the promising setting from industrial manufacturing and review current IAD algorithms in our proposed setting. In addition, we investigate in depth which network architecture design can considerably improve anomaly detection performance. In the end, we highlight several exciting future research directions for image anomaly detection.

Acknowledgements

This work was partly supported by the National Key R&D Program of China (No. 2022YFF1202903) and National Natural Science Foundation of China (Nos. 62122035 and 62206122). Y. Jin is funded by an Alexander von Humboldt Professorship for Artificial Intelligence endowed by the Federal Ministry of Education and Research of Germany. Open Access funding provided by Bielefeld University.

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

Open Access

References

[1] T. Czimmermann, G. Ciuti, M. Milazzo, M. Chiurazzi, S. Roccella, C. M. Oddo, P. Dario. Visual-based defect detection and classification approaches for industrial applications – A survey. *Sensors*, vol. 20, no. 5, Article number 1459, 2020. DOI: 10.3390/s20051459.

[2] X. Tao, X. Y. Gong, X. Zhang, S. H. Yan, C. Adak. Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Transactions on Instrumentation and Measurement*, vol. 71, Article number 5018021, 2022. DOI: 10.1109/TIM.2022.3196436.

[3] Y. J. Cui, Z. X. Liu, S. G. Lian. A survey on unsupervised industrial anomaly detection algorithms, [Online], Available: https://arxiv.org/abs/2204.11161, 2022.

[4] Z. Y. You, L. Cui, Y. J. Shen, K. Yang, X. Lu, Y. Zheng, X. Y. Le. A unified model for multi-class anomaly detection. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, New Orleans, USA, 2022.

[5] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger. MVTec ad — a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 9584–9592, 2019. DOI: 10.1109/CVPR.2019.00982.

[6] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, C. Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision*, vol. 130, no. 4, pp. 947–969, 2022. DOI: 10.1007/s11263-022-01578-9.

[7] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[8] K. Simonyan. A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, USA, 2015. DOI: 10.48550/arXiv.1409.1556.

[9] M. X. Tan, Q. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International conference on machine learning*, Long Beach, USA, pp. 6105–6114, 2019.

[10] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 4182–4191, 2020. DOI: 10.1109/CVPR42600.2020.00424.

[11] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, H. R. Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, USA, pp. 14897–14907, 2021. DOI: 10.1109/CVPR46437.2021.01466.

[12] G. D. Wang, S. M. Han, E. R. Ding, D. Huang. Student-teacher feature pyramid matching for anomaly detection. In *Proceedings of 32nd British Machine Vision Conference 2021*, 2021.

[13] S. Yamada, K. Hotta. Reconstruction student with attention for student-teacher pyramid matching, [Online], Available: https://arxiv.org/abs/2111.15376, 2021.

[14] H. Q. Deng, X. Y. Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 9727–9736, 2022. DOI: 10.1109/CVPR52688.2022.00951.

[15] Y. K. Cao, Q. Wan, W. M. Shen, L. Gao. Informative knowledge distillation for image anomaly segmentation. *Knowledge-based Systems*, vol. 248, Article number 108846, 2022. DOI: 10.1016/j.knosys.2022.108846.

[16] M. Rudolph, T. Wehrbein, B. Rosenhahn, B. Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 2591–2601, 2023. DOI: 10.1109/WACV56688.2023.00262.

[17] S. Yamada, S. Kamiya, K. Hotta. Reconstructed student-teacher and discriminative networks for anomaly detection. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Kyoto, Japan, pp. 2725–2732, 2022. DOI: 10.1109/IROS47612.2022.9981509.

[18] M. H. Yang, P. Wu, J. Liu, H. Feng. MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities, [Online], Available: https://arxiv.org/abs/2205.00908, 2022.

[19] J. H. Yi, S. Yoon. Patch SVDD: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the 15th Asian Conference on Computer Vision*, Kyoto, Japan, pp. 375–390, 2021. DOI: 10.1007/978-3-030-69544-6_23.

[20] Z. Zhang, X. G. Deng. Anomaly detection using improved deep SVDD model with data structure preservation. *Pattern Recognition Letters*, vol. 148, pp. 1–6, 2021. DOI: 10.1016/j.patrec.2021.04.020.

[21] C. F. Hu, K. Chen, H. Shao. A semantic-enhanced method based on deep SVDD for pixel-wise anomaly detection. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Shenzhen, China, 2021. DOI: 10.1109/ICME51207.2021.9428370.

[22] F. V. Massoli, F. Falchi, A. Kantarci, Ş. Akti, H. K. Ekenel, G. Amato. MOCCA: Multilayer one-class classification for anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2313–2323, 2022. DOI: 10.1109/TNNLS.2021.3130074.

[23] D. Sauter, A. Schmitz, F. Dikici, H. Baumgartl, R. Buettner. Defect detection of metal nuts applying convolutional neural networks. In *Proceedings of the 45th IEEE Annual Computers, Software, and Applications Conference*, Madrid, Spain, pp. 248–257, 2021. DOI: 10.1109/COMPSAC51774.2021.00043.

[24] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 1800–1807, 2017. DOI: 10.1109/CVPR.2017.195.

[25] T. Reiss, N. Cohen, L. Bergman, Y. Hoshen. PANDA: Adapting pretrained features for anomaly detection and segmentation. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, USA, pp. 2805–2813, 2021. DOI: 10.1109/CVPR46437.2021.00283.

[26] L. Bergman, N. Cohen, Y. Hoshen. Deep nearest neighbor anomaly detection, [Online], Available: https://arxiv.org/abs/2002.10445, 2020.

[27] K. Sohn, C. L. Li, J. Yoon, M. Jin, T. Pfister. Learning and evaluating representations for deep one-class classification. In *Proceedings of the 9th International Conference on Learning Representations*, 2021.

[28] X. L. Bai, Y. M. Fang, W. S. Lin, L. P. Wang, B. F. Ju. Saliency-based defect detection in industrial images by using phase spectrum. *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2135–2145, 2014. DOI: 10.1109/TII.2014.2359416.

[29] M. H. Niu, K. C. Song, L. M. Huang, Q. Wang, Y. H. Yan, Q. G. Meng. Unsupervised saliency detection of rail surface defects using stereoscopic images. *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2271–2281, 2021. DOI: 10.1109/TII.2020.3004397.

[30] Y. H. Qiu, L. X. Tang, B. Li, S. L. Niu, T. Z. Niu. Uneven illumination surface defects inspection based on saliency detection and intrinsic image decomposition. *IEEE Access*, vol. 8, pp. 190663–190676, 2020. DOI: 10.1109/ACCESS.2020.3032108.

[31] C. L. Li, K. Sohn, J. Yoon, T. Pfister. CutPaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, USA, pp. 9659–9669, 2021. DOI: 10.1109/CVPR46437.2021.00954.

[32] S. Yoa, S. Lee, C. Kim, H. J. Kim. Self-supervised learning for anomaly detection with dynamic local augmentation. *IEEE Access*, vol. 9, pp. 147201–147211, 2021. DOI: 10.1109/ACCESS.2021.3124525.

[33] P. de Haan, S. Löwe. Contrastive predictive coding for anomaly detection, [Online], Available: https://arxiv.org/abs/2107.07820, 2021.

[34] D. M. J. Tax, R. P. W. Duin. Support vector data description. *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004. DOI: 10.1023/B:MACH.0000008084.60811.49.

[35] P. Liznerski, L. Ruff, R. A. Vandermeulen, B. J. Franks, M. Kloft, K. R. Muller. Explainable deep one-class classification. In *Proceedings of the 9th International Conference on Learning Representations*, 2021.

[36] T. Reiss, Y. Hoshen. Mean-shifted contrastive loss for anomaly detection, [Online], Available: https://arxiv.org/abs/2106.03844, 2021.

[37] A. van den Oord, Y. Z. Li, O. Vinyals. Representation learning with contrastive predictive coding, [Online], Available: https://arxiv.org/abs/1807.03748, 2018.

[38] D. W. Zhang, J. W. Han, Y. Zhang, D. Xu. Synthesizing supervision for learning deep saliency network without human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 7, pp. 1755–1769, 2020. DOI: 10.1109/TPAMI.2019.2900649.

[39] M. C. Zhuge, D. P. Fan, N. Liu, D. W. Zhang, D. Xu, L. Shao. Salient object detection via integrity learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3738–3752, 2023. DOI: 10.1109/TPAMI.2022.3179526.

[40] C. W. Fang, H. B. Tian, D. W. Zhang, Q. Zhang, J. G. Han, J. W. Han. Densely nested top-down flows for salient object detection. *Science China Information Sciences*, vol. 65, no. 8, Article number 182103, 2022. DOI: 10.1007/s11432-021-3384-y.

[41] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 618–626, 2017. DOI: 10.1109/ICCV.2017.74.

[42] S. Venkataramanan, K. C. Peng, R. V. Singh. A. Mahalanobis. Attention guided anomaly localization in images. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 485–503, 2020. DOI: 10.1007/978-3-030-58520-4_29.

[43] A. S. Iquebal, S. T. Bukkapatnam. Consistent estimation of the max-flow problem: Towards unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2346–2357, 2022. DOI: 10.1109/TPAMI.2020.3039745.

[44] H. Attar, M. Calin, L. C. Zhang, S. Scudino, J. Eckert. Manufacture by selective laser melting and mechanical behavior of commercially pure titanium. *Materials Science and Engineering*: *A*, vol. 593, pp. 170–177, 2014. DOI: 10.1016/j.msea.2013.11.038.

[45] M. Tailanian, P. Musé, Á. Pardo. A multi-scale a contrario method for unsupervised image anomaly detection. In *Proceedings of the 20th IEEE International Conference on Machine Learning and Applications*, Pasadena, USA, pp. 179–184, 2021. DOI: 10.1109/ICMLA52953.2021.00035.

[46] O. Rippel, P. Mertens, D. Merhof. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In *Proceedings of the 25th International Conference on Pattern Recognition*, Milan, Italy, pp. 6726–6733, 2021. DOI: 10.1109/ICPR48806.2021.9412109.

[47] O. Rippel, A. Chavan, C. C. Lei, D. Merhof. Transfer learning gaussian anomaly detection by fine-tuning representations. In *Proceedings of the 2nd International Conference on Image Processing and Vision Engineering*, pp. 45–56, 2022. DOI: 10.5220/0011063900003209.

[48] K. T. Zhang, B. Wang, C. C. J. Kuo. PEDENet: Image anomaly localization via patch embedding and density estimation. *Pattern Recognition Letters*, vol. 153, pp. 144–150, 2022. DOI: 10.1016/j.patrec.2021.11.030.

[49] Q. Wan, L. Gao, X. Y. Li, L. Wen. Unsupervised image anomaly detection and segmentation based on pretrained feature mapping. *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2330–2339, 2023. DOI: 10.1109/TII.2022.3182385.

[50] Q. Wan, Y. K. Cao, L. Gao, W. M. Shen, X. Y. Li. Position encoding enhanced feature mapping for image anomaly detection. In *Proceedings of the 18th IEEE International Conference on Automation Science and Engineering*, Mexico City, Mexico, pp. 876–881, 2022. DOI: 10.1109/CASE49997.2022.9926547.

[51] Y. Zheng, X. Wang, R. Deng, T. P. Bao, R. Zhao, L. W. Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. In *Proceedings of IEEE International Conference on Multimedia and Expo*, Taipei, China, pp. 1–6, 2022. DOI: 10.1109/ICME52920.2022.9859925.

[52] M. Rudolph, B. Wandt, B. Rosenhahn. Same same but DifferNet: Semi-supervised defect detection with normalizing flows. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 1906–1915, 2021. DOI: 10.1109/WACV48630.2021.00195.

[53] M. Rudolph, T. Wehrbein, B. Rosenhahn, B. Wandt. Fully convolutional cross-scale-flows for image-based de-

fect detection. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 1829–1838, 2022. DOI: 10.1109/WACV51458.2022.00189.

[54] D. Gudovskiy, S. Ishizaka, K. Kozuka. CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 1819–1828, 2022. DOI: 10.1109/WACV51458.2022.00188.

[55] R. Q. Yan, F. Zhang, M. Y. Huang, W. Liu, D. Y. Hu, J. F. Li, Q. Liu, J. R. Jiang, Q. J. Guo, L. H. Zheng. CAIN-NFlow: Convolutional block attention modules and invertible neural networks flow for anomaly detection and localization tasks, [Online], Available: https://arxiv.org/abs/2206.01992, 2022.

[56] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby. An image is worth 16×16 words: Transformers for image recognition at scale. In *Proceedings of the 9th International Conference on Learning Representations*, 2020.

[57] J. W. Yu, Y. Zheng, X. Wang, W. Li, Y. S. Wu, R. Zhao, L. W. Wu. FastFlow: Unsupervised anomaly detection and localization via 2D normalizing flows, [Online], Available: https://arxiv.org/abs/2111.07677, 2021.

[58] Y. Kim, H. Jang, D. Lee, H. J. Choi. AltUB: Alternating training method to update base distribution of normalizing flow for anomaly detection, [Online], Available: https://arxiv.org/abs/2210.14913, 2022.

[59] D. J. Rezende, S. Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, pp. 1530–1538, 2015.

[60] Z. Liu, Y. T. Lin, Y. Cao, H. Hu, Y. X. Wei, Z. Zhang, S. Lin, B. N. Guo. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 9992–10002, 2021. DOI: 10.1109/ICCV48922.2021.00986.

[61] S. Woo, J. Park, J. Y. Lee, I. S. Kweon. CBAM: Convolutional block Attention module. In *Proceedings of the 15th European Conference on Computer Vision*, Springer, Munich, Germany, pp. 3–19, 2018. DOI: 10.1007/978-3-030-01234-2_1.

[62] J. H. Kim, D. H. Kim, S. Yi, T. Lee. Semi-orthogonal embedding for efficient unsupervised anomaly segmentation, [Online], Available: https://arxiv.org/abs/2105.14737, 2021.

[63] N. Cohen, Y. Hoshen. Sub-image anomaly detection with deep pyramid correspondences, [Online], Available: https://arxiv.org/abs/2005.02357, 2020.

[64] N. Li, K. T. Jiang, Z. H. Ma, X. Wei, X. P. Hong, Y. H. Gong. Anomaly detection via self-organizing map. In *Proceedings of the IEEE International Conference on Image Processing*, Anchorage, USA, pp. 974–978, 2021. DOI: 10.1109/ICIP42928.2021.9506433.

[65] Q. Wan, L. Gao, X. Y. Li, L. Wen. Industrial image anomaly localization based on gaussian clustering of pretrained feature. *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6182–6192, 2022. DOI: 10.1109/TIE.2021.3094452.

[66] C. C. Tsai, T. H. Wu, S. H. Lai. Multi-scale patch-based representation learning for image anomaly detection and segmentation. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 3065–3073, 2022. DOI: 10.1109/WACV51458.2022.00312.

[67] Y. Zou, J. Jeong, L. Pemula, D. Q. Zhang, O. Dabeer. SPot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Proceedings of the 17th European Conference on Computer Vision*, Springer, Tel Aviv, Israel, pp. 392–408, 2022. DOI: 10.1007/978-3-031-20056-4_23.

[68] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, P. Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 14298–14308, 2022. DOI: 10.1109/CVPR52688.2022.01392.

[69] S. Lee, S. Lee, B. C. Song. CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*, vol. 10, pp. 78446–78454, 2022. DOI: 10.1109/ACCESS.2022.3193699.

[70] D. Kim, C. Park, S. Cho, S. Lee. FAPM: Fast adaptive patch memory for real-time industrial anomaly detection, [Online], Available: https://arxiv.org/abs/2211.07381, 2022.

[71] J. Bae, J. H. Lee, S. Kim. Image anomaly detection and localization with position and neighborhood information, [Online], Available: https://arxiv.org/abs/2211.12634, 2022.

[72] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, S. Stolfo. *Applications of Data Mining in Computer Security*, D. Barbará, S. Jajodia, Eds., New York, USA: Springer, pp. 77–101, 2002. DOI: 10.1007/978-1-4615-0953-0_4.

[73] T. Defard, A. Setkov, A. Loesch, R. Audigier. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *Proceedings of Intemational Conference on Pattern Recognition – ICPR International Workshops and Challenges*, Springer, Milano, Italy, pp. 475–489, 2021. DOI: 10.1007/978-3-030-68799-1_35.

[74] J. Jang, E. Hwang, S. H. Park. N-pad: Neighboring pixel-based industrial anomaly detection, [Online], Available: https://arxiv.org/abs/2210.08768, 2022.

[75] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, C. Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Funchal, Portugal, pp. 372–380, 2018. DOI: 10.5220/0007364503720380.

[76] H. Chung, J. Park, J. Keum, H. Ki, S. Kang. Unsupervised anomaly detection using style distillation. *IEEE Access*, vol. 8, pp. 221494–221502, 2020. DOI: 10.1109/ACCESS.2020.3043473.

[77] Y. F. Liu, C. Q. Zhuang, F. Lu. Unsupervised two-stage anomaly detection, [Online], Available: https://arxiv.org/abs/2103.11671, 2021.

[78] J. Yang, Y. Shi, Z. Q. Qi. DFR: Deep feature reconstruction for unsupervised anomaly segmentation, [Online], Available: https://arxiv.org/abs/2012.07122, 2020.

[79] Y. Yan, D. M. Wang, G. L. Zhou, Q. J. Chen. Unsupervised anomaly segmentation via multilevel image reconstruction and adaptive attention-level transition. *IEEE Transactions on Instrumentation and Measurement*, vol. 70, Article number 5015712, 2021. DOI: 10.1109/TIM.2021.3107586.

[80] K. Zhou, Y. T. Xiao, J. L. Yang, J. Cheng, W. Liu, W. X. Luo, Z. W. Gu, J. Liu, S. H. Gao. Encoding structure-tex-

ture relation with p-net for anomaly detection in retinal images. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 360–377, 2020. DOI: 10.1007/978-3-030-58565-5_22.

[81] A. S. Collin, C. De Vleeschouwer. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In *Proceedings of the 25th International Conference on Pattern Recognition*, Milan, Italy, pp. 7915–7922, 2021. DOI: 10.1109/ICPR48806.2021.9412842.

[82] X. Tao, D. P. Zhang, W. Z. Ma, Z. X. Hou, Z. F. Lu, C. Adak. Unsupervised anomaly detection for surface defects with dual-siamese network. *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7707–7717, 2022. DOI: 10.1109/TII.2022.3142326.

[83] J. L. Hou, Y. Y. Zhang, Q. Y. Zhong, D. Xie, S. L. Pu, H. Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 8771–8780, 2021. DOI: 10.1109/ICCV48922.2021.00867.

[84] T. K. Liu, B. Li, Z. Zhao, X. Du, B. K. Jiang, L. Q. Geng. Reconstruction from edge image combined with color and gradient difference for industrial surface anomaly detection, [Online], Available: https://arxiv.org/abs/2210.14485, 2022.

[85] D. Kim, D. Jeong, H. Kim, K. Chong, S. Kim, H. Cho. Spatial contrastive learning for anomaly detection and localization. *IEEE Access*, vol. 10, pp. 17366–17376, 2022. DOI: 10.1109/ACCESS.2022.3149130.

[86] Z. Y. Li, N. Li, K. T. Jiang, Z. H. Ma, X. Wei, X. P. Hong, Y. H. Gong. Superpixel masking and inpainting for self-supervised anomaly detection. In *Proceedings of the 31st British Machine Vision Conference 2020*, BMVC, 2020.

[87] V. Zavrtanik, M. Kristan, D. Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, vol. 112, Article number 107706, 2021. DOI: 10.1016/j.patcog.2020.107706.

[88] H. Nakanishi, M. Suzuki, Y. Matsuo. Iterative image inpainting with structural similarity mask for anomaly detection. In *Proceedings of the 9th International Conference on Learning Representations*, 2021.

[89] A. Bauer. Self-supervised training with autoencoders for visual anomaly detection, [Online], Available: https://arxiv.org/abs/2206.11723, 2022.

[90] C. Q. Huang, Q. W. Xu, Y. F. Wang, Y. Wang, Y. Zhang. Self-supervised masking for unsupervised anomaly detection and localization. *IEEE Transactions on Multimedia*, to be published.

[91] V. Zavrtanik, M. Kristan, D. Skočaj. DRAEM – a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 8310–8319, 2021. DOI: 10.1109/ICCV48922.2021.00822.

[92] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Munich, Germany, pp. 234–241, 2015. DOI: 10.1007/978-3-319-24574-4_28.

[93] P. Xing, Y. P. Sun, Z. C. Li. Self-supervised guided segmentation framework for unsupervised anomaly detection, [Online], Available: https://arxiv.org/abs/2209.12440, 2022.

[94] V. Zavrtanik, M. Kristan, D. Skočaj. DSR – A dual subspace re-projection network for surface anomaly detection. In *Proceedings of the 17th European Conference on Computer Vision*, Springer, Tel Aviv, Israel, pp. 539–554, 2022. DOI: 10.1007/978-3-031-19821-2_31.

[95] H. M. Schlüter, J. Tan, B. Hou, B. Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *Proceedings of the 17th European Conference on Computer Vision*, Springer, Tel Aviv, Israel, pp. 474–489, 2022. DOI: 10.1007/978-3-031-19821-2_27.

[96] N. C. Ristea, N. Madan, R. T. Ionescu, K. Nasrollahi, F. S. Khan, T. B. Moeslund, M. Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 13566–13576, 2022. DOI: 10.1109/CVPR52688.2022.01321.

[97] N. Madan, N. C. Ristea, R. T. Ionescu, K. Nasrollahi, F. S. Khan, T. B. Moeslund, M. Shah. Self-supervised masked convolutional transformer block for anomaly detection, [Online], Available: https://arxiv.org/abs/2209.12148, 2022.

[98] D. Dehaene, O. Frigo, S. Combrexelle, P. Eline. Iterative energy-based projection on a normal data manifold for anomaly localization. In *Proceedings of the 8th International Conference on Learning Representations*, Addis Ababa, Ethiopia, 2020.

[99] W. Q. Liu, R. Z. Li, M. Zheng, S. Karanam, Z. Y. Wu, B. Bhanu, R. J. Radke, O. Camps. Towards visually explaining variational autoencoders. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 8639–8648, 2020. DOI: 10.1109/CVPR42600.2020.00867.

[100] T. Matsubara, K. Sato, K. Hama, R. Tachibana, K. Uehara. Deep generative model using unregularized score for anomaly detection with heterogeneous complexity. *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5161–5173, 2022. DOI: 10.1109/TCYB.2020.3027724.

[101] D. Dehaene, P. Eline. Anomaly localization by modeling perceptual features, [Online], Available: https://arxiv.org/abs/2008.05369, 2020.

[102] L. Wang, D. K. Zhang, J. H. Guo, Y. X. Han. Image anomaly detection using normal data only by latent space resampling. *Applied Sciences*, vol. 10, no. 23, Article number 8660, 2020. DOI: 10.3390/app10238660.

[103] X. D. Yan, H. D. Zhang, X. X. Xu, X. Hu, P. A. Heng. Learning semantic context from normal samples for unsupervised anomaly detection. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, pp. 3110–3118, 2021. DOI: 10.1609/aaai.v35i4.16420.

[104] J. Song, K. Kong, Y. I. Park, S. G. Kim, S. J. Kang. AnoSeg: Anomaly segmentation network using self-supervised learning, [Online], Available: https://arxiv.org/abs/2110.03396, 2021.

[105] Y. F. Liang, J. N. Zhang, S. W. Zhao, R. Z. Wu, Y. Liu, S. W. Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection, [Online], Available: https://arxiv.org/abs/2203.00259, 2022.

[106] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, G. L. Foresti. VT-ADL: A vision transformer network for image anomaly detection and localization. In *Proceedings of the IEEE 30th International Symposium on Industrial Electronics*, Kyoto, Japan, 2021. DOI: 10.1109/ISIE45552.2021.9576231.

[107] Z. Y. You, K. Yang, W. H. Luo, L. Cui, Y. Zheng, X. Y. Le. ADTR: Anomaly detection transformer with feature reconstruction. In *Proceedings of the 29th International Conference on Neural Information Processing*, Springer, pp. 298–310, 2023. DOI: 10.1007/978-3-031-30111-7_26.

[108] Y. Lee, P. Kang. AnoViT: Unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. *IEEE Access*, vol. 10, pp. 46717–46724, 2022. DOI: 10.1109/ACCESS.2022.3171559.

[109] E. Mathian, H. Liu, L. Fernandez-Cuesta, D. Samaras, M. Foll, L. Chen. HaloAE: An HaloNet based local transformer auto-encoder for anomaly detection and localization, [Online], Available: https://arxiv.org/abs/2208.03486, 2022.

[110] J. Pirnay, K. Chai. Inpainting transformer for anomaly detection. In *Proceedings of the 21st International Conference on Image Analysis and Processing*, Springer, Lecce, Italy, pp. 394–406, 2022. DOI: 10.1007/978-3-031-06430-2_33.

[111] J. L. Jiang, J. L. Zhu, M. Bilal, Y. Cui, N. Kumar, R. H. Dou, F. Su, X. L. Xu. Masked Swin Transformer Unet for industrial anomaly detection. *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2200–2209, 2023. DOI: 10.1109/TII.2022.3199228.

[112] A. De Nardin, P. Mishra, G. L. Foresti, C. Piciarelli. Masked transformer for image anomaly localization. *International Journal of Neural Systems*, vol. 32, no. 7, Article number 2250030, 2022. DOI: 10.1142/S0129065722500307.

[113] J. Wyatt, A. Leach, S. M. Schmon, C. G. Willcocks. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 649–655, 2022. DOI: 10.1109/CVPRW56347.2022.00080.

[114] Y. P. Teng, H. Y. Li, F. Z. Cai, M. Shao, S. Y. Xia. Unsupervised visual defect detection with score-based generative model, [Online], Available: https://arxiv.org/abs/2211.16092, 2022.

[115] J. Hu, L. Shen, G. Sun. Squeeze-and-excitation networks. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 7132–7141, 2018. DOI: 10.1109/CVPR.2018.00745.

[116] A. Vaswani, P. Ramachandran, A. Srinivas, N. Parmar, B. Hechtman, J. Shlens. Scaling local self-attention for parameter efficient visual backbones. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, USA, pp. 12889–12899, 2021. DOI: 10.1109/CVPR46437.2021.01270.

[117] K. Perlin. An image synthesizer. *ACM SIGGRAPH Computer Graphics*, vol. 19, no. 3, pp. 287–296, 1985. DOI: 10.1145/325165.325247.

[118] J. Ho, A. Jain. P. Abbeel, Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp. 574, 2020.

[119] W. H. Chu, K. M. Kitani. Neural batch sampling with reinforcement learning for semi-supervised anomaly detection. In *Proceedings of the 16th European Conference on Computer Vision*, Springer, Glasgow, UK, pp. 751–766, 2020. DOI: 10.1007/978-3-030-58574-7_45.

[120] J. Božič, D. Tabernik, D. Skočaj. Mixed supervision for surface-defect detection: From weakly to fully supervised learning. *Computers in Industry*, vol. 129, Article number 103459, 2021. DOI: 10.1016/j.compind.2021.103459.

[121] G. S. Pang, C. B. Ding, C. H. Shen, A. van den Hengel. Explainable deep few-shot anomaly detection with deviation networks, [Online], Available: https://arxiv.org/abs/2108.00462, 2021.

[122] Q. Wan, L. Gao, X. Y. Li. Logit inducing with abnormality capturing for semi-supervised image anomaly detection. *IEEE Transactions on Instrumentation and Measurement*, vol. 71, Article number 3523412, 2022. DOI: 10.1109/TIM.2022.3205674.

[123] C. B. Ding, G. S. Pang, C. H. Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 7378–7388, 2022. DOI: 10.1109/CVPR52688.2022.00724.

[124] V. A. Sindagi, S. Srivastava. Domain adaptation for automatic OLED panel defect detection using adaptive support vector data description. *International Journal of Computer Vision*, vol. 122, no. 2, pp. 193–211, 2017. DOI: 10.1007/s11263-016-0953-y.

[125] K. P. Qiu, L. Tian, P. Wang. An effective framework of automated visual surface defect detection for metal parts. *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20412–20420, 2021. DOI: 10.1109/JSEN.2021.3095410.

[126] G. Bhattacharya, B. Mandal, N. B. Puhan. Interleaved deep artifacts-aware attention mechanism for concrete structural defect classification. *IEEE Transactions on Image Processing*, vol. 30, pp. 6957–6969, 2021. DOI: 10.1109/TIP.2021.3100556.

[127] Z. Y. Zeng, B. Liu, J. L. Fu, H. Y. Chao. Reference-based defect detection network. *IEEE Transactions on Image Processing*, vol. 30, pp. 6637–6647, 2021. DOI: 10.1109/TIP.2021.3096067.

[128] G. R. Song, K. C. Song, Y. H. Yan. Saliency detection for strip steel surface defects using multiple constraints and improved texture features. *Optics and Lasers in Engineering*, vol. 128, Article number 106000, 2020. DOI: 10.1016/j.optlaseng.2019.106000.

[129] X. M. Long, B. Fang, Y. F. Zhang, G. Y. Luo, F. C. Sun. Fabric defect detection using tactile information. In *Proceedings of IEEE International Conference on Robotics and Automation*, Xi'an, China, pp. 11169–11174, 2021. DOI: 10.1109/ICRA48506.2021.9561092.

[130] B. Z. Hu, B. Gao, W. L. Woo, L. F. Ruan, J. K. Jin, Y. Yang, Y. J. Yu. A lightweight spatial and temporal multi-feature fusion network for defect detection. *IEEE Transactions on Image Processing*, vol. 30, pp. 472–486, 2021. DOI: 10.1109/TIP.2020.3036770.

[131] M. Ferguson, R. Ak, Y. T. T. Lee, K. H. Law. Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning, [Online], Available: https://arxiv.org/abs/1808.02518, 2018.

[132] X. Tao, D. P. Zhang, W. Z. Ma, X. L. Liu, D. Xu. Automatic metallic surface defect detection and recognition with convolutional neural networks. *Applied Sciences*, vol. 8, no. 9, Article number 1575, 2018. DOI: 10.3390/app8091575.

[133] J. Y. Li, Z. F. Su, J. H. Geng, Y. X. Yin. Real-time detection of steel strip surface defects based on improved YOLO detection network. *IFAC-PapersOnLine*, vol. 51, no. 21, pp. 76–81, 2018. DOI: 10.1016/j.ifacol.2018.09.412.

[134] D. Tabernik, S. Šela, J. Skvarč, D. Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, vol. 31, no. 3, pp. 759–776, 2020. DOI: 10.1007/s10845-019-01476-x.

[135] D. W. Zhang, G. Y. Guo, W. Y. Zeng, L. Li, J. W. Han. Generalized weakly supervised object localization. *IEEE Transactions on Neural Networks and Learning Systems*, to be published.

[136] D. W. Zhang, W. Y. Zeng, J. R. Yao, J. W. Han. Weakly supervised object detection using proposal- and semantic-level relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3349–3363, 2022. DOI: 10.1109/TPAMI.2020.3046647.

[137] P. L. Huang, J. W. Han, N. Liu, J. Ren, D. W. Zhang. Scribble-supervised video object segmentation. *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 2, pp. 339–353, 2022. DOI: 10.1109/JAS.2021.1004210.

[138] Y. Xi, K. Zhou, L. W. Meng, B. Chen, H. M. Chen, J. Y. Zhang. Transmission Line Insulator Defect Detection Based on Swin Transformer and Context. *Machine Intelligence Research*, vol. 20, no. 5, pp. 729–740, 2023. DOI: 10.1007/s11633-022-1355-y.

[139] J. C. Wu, D. J. Chen, C. S. Fuh, T. L. Liu. Learning unsupervised metaformer for anomaly detection. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 4349–4358, 2021. DOI: 10.1109/ICCV48922.2021.00433.

[140] A. M. Kamoona, A. K. Gostar, A. Bab-Hadiashar, R. Hoseinnezhad. Anomaly detection of defect using energy of point pattern features within random finite set framework, [Online], Available: https://arxiv.org/abs/2108.12159, 2021.

[141] C. Q. Huang, H. Y. Guan, A. F. Jiang, Y. Zhang, M. Spratling, Y. F. Wang. Registration based few-shot anomaly detection. In *Proceedings of the 17th European Conference on Computer Vision*, Springer, Tel Aviv, Israel, pp. 303–319, 2022. DOI: 10.1007/978-3-031-20053-3_18.

[142] E. Schwartz, A. Arbelle, L. Karlinsky, S. Harary, F. Scheidegger, S. Doveh, R. Giryes. MAEDAY: MAE for few and zero shot AnomalY-detection, [Online], Available: https://arxiv.org/abs/2211.14307, 2022.

[143] K. M. He, X. L. Chen, S. N. Xie, Y. H. Li, P. Dollár, R. Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, USA, pp. 15979–15988, 2022. DOI: 10.1109/CVPR52688.2022.01553.

[144] J. Jeong, Y. Zou, T. Kim, D. Q. Zhang, A. Ravichandran, O. Dabeer. WinCLIP: Zero-/few-shot anomaly classification and segmentation, [Online], Available: https://arxiv.org/abs/2303.14814, 2023.

[145] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, I. Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8748–8763, 2021.

[146] D. S. Tan, Y. C. Chen, T. P. C. Chen, W. C. Chen. TrustMAE: A noise-resilient defect classification framework using memory-augmented auto-encoders with trust regions. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 276–285, 2021. DOI: 10.1109/WACV48630.2021.00032.

[147] J. Yoon, K. Sohn, C. L. Li, S. Ö. Arik, C. Y. Lee, T. Pfister. Self-supervise, refine, repeat: Improving unsupervised anomaly detection. *Transactions on Machine Learning Research*, to be published.

[148] C. Qiu, A. D. Li, M. Kloft, M. Rudolph, S. Mandt. Latent outlier exposure for anomaly detection with contaminated data. In *Proceedings of the 39th International Conference on Machine Learning*, Baltimore, USA, pp. 18153–18167, 2022.

[149] Y. H. Chen, Y. Tian, G. S. Pang, G. Carneiro. Deep one-class classification via interpolated gaussian descriptor. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 383–392, 2022. DOI: 10.1609/aaai.v36i1.19915.

[150] P. Bergmann, D. Sattlegger. Anomaly detection in 3D point clouds using deep geometric descriptors. In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 2612–2622, 2023. DOI: 10.1109/WACV56688.2023.00264.

[151] E. Horwitz, Y. Hoshen. Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Vancouver, Canada, pp. 2967–2976, 2023.

[152] R. B. Rusu, N. Blodow, M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In *Proceedings of IEEE International Conference on Robotics and Automation*, Kobe, Japan, pp. 3212–3217, 2009. DOI: 10.1109/ROBOT.2009.5152473.

[153] T. Reiss, N. Cohen, E. Horwitz, R. Abutbul, Y. Hoshen. Anomaly detection requires better representations. In *Proceedings of Computer Vision – 2022 Workshops*, Springer, Tel Aviv, Israel, pp. 56–68, 2023. DOI: 10.1007/978-3-031-25069-9_4.

[154] J. H. Liu, C. Y. Wang, H. Su, B. Du, D. C. Tao. Multistage GAN for fabric defect detection. *IEEE Transactions on Image Processing*, vol. 29, pp. 3388–3400, 2020. DOI: 10.1109/TIP.2019.2959741.

[155] O. Rippel, M. Müller, D. Merhof, GAN-based defect synthesis for anomaly detection in fabrics. In *Proceedings of the 25th IEEE International Conference on Emerging Technologies and Factory Automation*, Vienna, Austria, pp. 534–540, 2020. DOI: 10.1109/ETFA46521.2020.9212099.

[156] J. Y. Zhu, T. Park, P. Isola, A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 2242–2251, 2017. DOI: 10.1109/ICCV.2017.244.

[157] S. L. Niu, B. Li, X. G. Wang, H. Lin. Defect image sample generation with GAN for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 3, pp. 1611–1622, 2020. DOI: 10.1109/tase.2020.2967415.

[158] T. R. Wei, D. H. Cao, X. R. Jiang, C. Y. Zheng, L. Z. Liu. Defective samples simulation through neural style transfer for automatic surface defect segment. In *Proceedings of International Conference on Optical Instruments and Technology: Optoelectronic Measurement Technology and Systems*, Beijing, China, pp. 1143904, 2020. DOI: 10.1117/12.2540464.

[159] T. R. Wei, D. H. Cao, C. Y. Zheng, Q. Yang. A simulation-based few samples learning method for surface defect segmentation. *Neurocomputing*, vol. 412, pp. 461–476, 2020. DOI: 10.1016/j.neucom.2020.06.090.

[160] S. Jain, G. Seth, A. Paruthi, U. Soni, G. Kumar. Synthetic data augmentation for surface defect detection and classification using deep learning. *Journal of Intelligent Manufacturing*, vol. 33, no. 4, pp. 1007–1020, 2022. DOI: 10.1007/s10845-020-01710-x.

[161] R. Y. Wang, S. Hoppe, E. Monari, M. F. Huber. Defect transfer GAN: Diverse defect synthesis for data augmentation. In *Proceedings of the 33rd British Machine Vision Conference 2022*, London, UK, pp. 445, 2021.

[162] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, pp. 6629–6640, 2017.

[163] M. Bińkowski, D. J. Sutherland, M. Arbel, A. Gretton. Demystifying MMD GANs. In *Proceedings of the 6th International Conference on Learning Representations*, Vancouver, Canada, 2018.

[164] G. J. Zhang, K. W. Cui, T. Y. Hung, S. J. Lu. Defect-GAN: High-fidelity defect synthesis for automated defect inspection. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, USA, pp. 2523–2533, 2021. DOI: 10.1109/WACV48630.2021.00257.

[165] J. Silvestre-Blanes, T. Albero-Albero, I. Miralles, R. Pérez-Llorens, J. MoreA public fabric database for defect detection methods and results. *Autex Research Journal*, vol. 19, no. 4, pp. 363–374, 2019. DOI: 10.2478/aut-2019-0035.

[166] M. Wieler, T. Hahn. Weakly supervised learning for industrial optical inspection. *DAGM Symposium*, 2007.

[167] S. L. Tang, F. He, X. L. Huang, J. Yang. Online PCB defect detector on a new PCB defect dataset, [Online], Available: https://arxiv.org/abs/1902.06197, 2019.

[168] L. Bonfiglioli, M. Toschi, D. Silvestri, N. Fioraio, D. De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the 16th Asian Conference on Computer Vision*, Springer, Macao, China, pp. 459–475, 2022. DOI: 10.1007/978-3-031-26348-4_27.

[169] C. S. C. Tsang, H. Y. T. Ngan, G. K. H. Pang. Fabric inspection based on the elo rating method. *Pattern Recognition*, vol. 51, pp. 378–394, 2016. DOI: 10.1016/j.patcog.2015.09.022.

[170] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, M. Carrasco. Gdxray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, vol. 34, no. 4, Article number 42, 2015. DOI: 10.1007/s10921-015-0315-7.

[171] T. P. Bao, J. D. Chen, W. Li, X. Wang, J. J. Fei, L. W. Wu, R. Zhao, Y. Zheng. MIAD: A maintenance inspection dataset for unsupervised anomaly detection, [Online], Available: https://arxiv.org/abs/2211.13968, 2022.

[172] S. Jezek, M. Jonak, R. Burget, P. Dvorak, M. Skotak. Deep learning-based defect detection of metal parts: Evaluating current methods in complex conditions. In *Proceedings of the 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, Brno, Czech Republic, pp. 66–71, 2021. DOI: 10.1109/ICUMT54235.2021.9631567.

[173] Y. B. Huang, C. Y. Qiu, K. Yuan. Surface defect saliency of magnetic tile. *The Visual Computer*, vol. 36, no. 1, pp. 85–96, 2020. DOI: 10.1007/s00371-018-1588-5.

[174] P. Bergmann, X. Jin, D. Sattlegger, C. Steger. The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 202–213, 2021. DOI: 10.5220/0010865000003124.

[175] D. Carrera, F. Manganini, G. Boracchi, E. Lanzarone. Defect detection in SEM images of nanofibrous materials. *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 551–561, 2017. DOI: 10.1109/tii.2016.2641472.

[176] K. C. Song, Y. H. Yan. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, vol. 285, pp. 858–864, 2013. DOI: 10.1016/j.apsusc.2013.09.002.

[177] J. R. Gan, Q. Y. Li, J. Z. Wang, H. M. Yu. A hierarchical extractor-based visual rail surface inspection system. *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7935–7944, 2017. DOI: 10.1109/JSEN.2017.2761858.

[178] Kaggle. Severstal: Steel defect detection, [Online], Available: https://www.kaggle.com/c/severstal-steel-defect-detection, 2019.

[179] D. M. Yang, Y. R. Cui, Z. Y. Yu, H. Q. Yuan. Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 1237–1249, 2021. DOI: 10.1080/08839514.2021.1975391.

[180] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, C. Steger. The mvtec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1038–1059, 2021. DOI: 10.1007/s11263-020-01400-4.

[181] M. A. Rahman, Y. Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In *Proceedings of the 12th International Symposium on Advances in Visual Computing*, Springer, Las Vegas, USA, pp. 234–244, 2016. DOI: 10.1007/978-3-319-50835-1_22.

[182] S. Sheynin, S. Benaim, L. Wolf. A hierarchical transformation-discriminating generative model for few shot anomaly detection. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 8475–8484, 2021. DOI: 10.1109/ICCV48922.2021.00838.

[183] X. Jiang, J. L. Liu, J. B. Wang, Q. Nie, K. Wu, Y. Liu, C. J. Wang, F. Zheng. SoftPatch: Unsupervised anomaly detection with noisy data. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, New Orleans, USA, 2022.

[184] G. Y. Xie, J. B. Wang, J. Q. Liu, J. Y. Lyu, Y. Liu, C. J. Wang, F. Zheng, Y. C. Jin. IM-IAD: Industrial image anomaly detection benchmark in manufacturing, [Online], Available: https://arxiv.org/abs/2301.13359, 2023.

[185] D. W. Zhang, G. H. Huang, Q. Zhang, J. G. Han, J. W. Han, Y. Z. Wang, Y. Z. Yu. Exploring task structure for brain tumor segmentation from multi-modality MR images. *IEEE Transactions on Image Processing*, vol. 29, pp. 9032–9043, 2020. DOI: 10.1109/TIP.2020.3023609.

[186] D. W. Zhang, G. H. Huang, Q. Zhang, J. G. Han, J. W. Han, Y. Z. Yu. Cross-modality deep feature learning for brain tumor segmentation. *Pattern Recognition*, vol. 110, Article number 107562, 2021. DOI: 10.1016/j.patcog.2020.107562.

[187] C. W. Fang, Q. Wang, L. C. Cheng, Z. F. Gao, C. W. Pan, Z. Cao, Z. H. Zheng, D. W. Zhang. Reliable mutual distillation for medical image segmentation under imperfect annotations. *IEEE Transactions on Medical Imaging*, vol. 42, no. 6, pp. 1720–1734, 2023. DOI: 10.1109/TMI.2023.3237183.

[188] W. C. Wang, E. Ahn, D. Feng, J. Kim. A Review of Predictive and Contrastive Self-supervised Learning for Med-

ical Images. *Machine Intelligence Research*, vol. 20, no. 4, pp. 483–513, 2023. DOI: 10.1007/s11633-022-1406-4.

[189] T. Shyamalee, D. Meedeniya. Glaucoma Detection with Retinal Fundus Images Using Segmentation and Classification. *Machine Intelligence Research*, vol. 19, no. 6, pp. 563–580, 2022. DOI: 10.1007/s11633-022-1354-z.

Jiaqi Liu received the B. Sc. degree in software engineering from Dalian University of Technology, China in 2019. He is a master student in electronic science and technology from Southern University of Science and Technology, China, under the supervision of Professor Feng Zheng.

His research interest is image anomaly detection.

E-mail: liujq32021@mail.sustech.edu.cn
ORCID iD: 0000-0002-2153-8411

Guoyang Xie received the B. Sc. degree in physical electronic and the M. Sc. degree in robotics from University of Electronic Science and Technology of China, Hong Kong University of Science and Technology, China in 2009 and 2013, respectively. He is a Ph. D. degree candidate in machine learning from University of Surrey, UK. Prior to that, he was the Principle Perception Algorithm Engineer in Baidu and GAC, respectively.

His research interests include anomaly detection, medical imaging, neural architecture search and federated learning.

E-mail: guoyang.xie@ieee.org
ORCID iD: 0000-0001-8433-8153

Jinbao Wang received the Ph. D. degree in computer application technology from University of Chinese Academy of Sciences (UCAS), China in 2019. He is currently a research assistant professor with Southern University of Science and Technology (SUSTech), China.

His research interests include machine learning, computer vision, image anomaly detection, and graph representation learning.

E-mail: linkingring@163.com
ORCID iD: 0000-0001-5916-8965

Shangnian Li received the B. Sc. in measurement and control technology and instruments and the M. Sc. degree in computer science and technology from Huaiyin Institute of Technology, Beijing Union University, China in 2012 and 2016, respectively. He is currently the research assistant of Sustech VIP Lab, China. Prior to that, he was the vehicle networking engineer in Beijing Xiangzhi Technology Co., Ltd., China.

His research interests include internet of things and anomaly detection.

E-mail: lisn3@mail.sustech.edu.cn

Chengjie Wang received the B. Sc. degree in computer science from Shanghai Jiao Tong University, China in 2011, and double M. Sc. degrees in computer science from Shanghai Jiao Tong University, and Waseda University, Japan, in 2014. He is currently the Research Director of YouTu Lab, Tencent, China. And he is pursuing the Ph.D. degree in computer science at Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. He has authored or coauthored more than 90 refereed papers on major Computer Vision and Artificial Intelligence Conferences, such as CVPR, ICCV, ECCV, AAAI, IJCAI, and NeurIPS, and holds more than 100 patents in his research areas.

His research interests include computer vision and machine learning.

E-mail: jasoncjwang@tencent.com

Feng Zheng received the Ph.D. degree in electronic science and engineering from the University of Sheffield, UK in 2017. He is currently an assistant professor with Department of Computer Science and Engineering, Southern University of Science and Technology, China.

His research interests include machine learning, computer vision, and human-computer interaction.

E-mail: f.zheng@ieee.com (Corresponding author)
ORCID iD: 0000-0002-1701-9141

Yaochu Jin received the B. Sc., M. Sc. and Ph. D. degrees in automatic control from Zhejiang University, China in 1988, 1991 and 1996, respectively, and the Dr.Ing. degree in computer engineering from Ruhr University Bochum, Germany in 2001. He is presently an Alexander von Humboldt Professor for Artificial Intelligence endowed by the German Federal Ministry of Education and Research, Chair of Nature Inspired Computing and Engineering, Faculty of Technology, Bielefeld University, Germany. He is also a Distinguished Chair, Professor in Computational Intelligence, Department of Computer Science, University of Surrey, Guildford, UK. He was a "Finland Distinguished Professor" of University of Jyvskyl awarded by the Academy of Science and Finnish Funding Agency for Innovation, Finland, and "Changjiang Distinguished Visiting Professor" of Northeastern University, awarded by the Ministry of Education, China. Prof. Jin is the President-Elect of the IEEE Computational Intelligence Society and the Editor-in-Chief of *Complex & Intelligent Systems*. He was named by Clarivate as a "Highly Cited Researcher" from 2019 to 2022 consecutively. He is a Member of Academia Europaea and Fellow of IEEE.

His research interests include human-centered learning and optimization, synergies between evolution and learning, and evolutionary developmental artificial intelligence.

E-mail: yaochu.jin@uni-bielefeld.de (Corresponding author)
ORCID iD: 0000-0003-1100-0631