Neurocomputing 493 (2022) 497–535

Contents lists available at ScienceDirect

Neurocomputing

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

Survey paper

GAN-based anomaly detection: A review

Xuan Xia [a], Xizhou Pan [a], Nan Li [a], Xing He [a], Lin Ma [a,*], Xiaoguang Zhang [a], Ning Ding [a,b,*]

[a] Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518129, PR China
[b] Institute of Robotics and Intelligent Manufacturing, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, PR China

ARTICLE INFO

Article history:
Received 30 June 2021
Revised 20 December 2021
Accepted 27 December 2021
Available online 3 January 2022

Keywords:
Deep learning
Generative adversarial nets
Anomaly detection
Adversarial learning and inference
Representation learning

ABSTRACT

Supervised learning algorithms have shown limited use in the field of anomaly detection due to the unpredictability and difficulty in acquiring abnormal samples. In recent years, unsupervised or semi-supervised anomaly-detection algorithms have become more widely used in anomaly-detection tasks. As a form of unsupervised learning algorithm, generative adversarial networks (GAN/GANs) have been widely used in anomaly detection because GAN can make abnormal inferences using adversarial learning of the representation of samples. To provide inspiration for the research of GAN-based anomaly detection, this review reconsiders the concept of anomaly, provides three criteria for discussing the anomaly detection task, and discusses the current challenges of anomaly detection. For the existing works, this review focuses on the theoretical and technological evolution, theoretical basis, applicable tasks, and practical application of GAN-based anomaly detection. This review also addresses the current internal and external outstanding issues encountered by GAN-based anomaly detection and predicts and analyzes several future research directions in detail. This review summarizes more than 330 references related to GAN-based anomaly detection and provides detailed technical information for researchers who are interested in GANs and want to apply them to anomaly-detection tasks.

1. Introduction

Anomaly detection [1] refers to the task of identifying abnormal data that are significantly different from the majority of instances and has many important applications, including industrial product defect detection, infrastructure distress detection, and medical diagnosis. There are many reasons or causes for anomalies, including system failures, human errors, malicious operations, or natural environmental changes. Anomaly detection is an important tool, because, on the one hand, it can reduce the risk and cost, and on the other hand, anomalies can convey valuable information, such as the failure of critical parts of a system or the destruction of infrastructure. Therefore, anomaly detection is generally considered an essential step in various decision systems.

Deep learning methods, such as Convolution Neural Networks (CNN) [2], Recurrent Neural Networks (RNN) [3], Auto-Encoders (AE) [4] and GAN/GANs [5], have been applied to various anomaly detection tasks and achieved good performance. As the data being processed by these methods becomes larger and more complex, more deep learning algorithms have been proposed to deal with these complex data for anomaly detection. However, there is usually a large number of normal samples that have no (or few) abnormal samples, which often suffer from class imbalance problems [6]. At present, the performance of deep anomaly detection (DAD) methods often depends on extensive training samples; therefore, data imbalance is an impediment to their application.

Due to the ability to perform distribution fitting, generative models have become some of the best methods of anomaly detection. Among them, the most representative are the Variational AutoEncoder (VAE) [7] and GAN (including their variants).The goal of VAE is to minimize the lower bound of the logarithmic likelihood of the data, whereas the goal of GAN is to achieve a balance between the generator and the discriminator [8]. GAN is empirically known to generate higher quality and higher definition results than does VAE [9]; it has thus received considerable attention since it was proposed.

GAN shows superior ability in generating real image instances. By training with a dataset that contains only normal samples and learning the feature representations in latent space, the abnormal samples, which are poorly reconstructed, can be detected. The ability of GAN to generate data alleviates the problem of insufficient abnormal samples to a certain extent [10]. Additionally, GAN is suitable for anomaly-detection tasks pertaining to complex datasets and can model high-dimensional data distributions. Further,

* Corresponding authors at: Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518129, PR China (N. Ding).
E-mail addresses: malin@cuhk.edu.cn (L. Ma), dingning@cuhk.edu.cn (N. Ding).

https://doi.org/10.1016/j.neucom.2021.12.093
0925-2312/

it has demonstrated state-of-the-art (SOTA) performance. Therefore, most current models and theories based on GAN are suitable for anomaly detection.

At present, GAN is an extremely popular research topic, and thus there are many research approaches on GAN-based anomaly detection that have been widely applied in industry, infrastructure, medical disease, and other areas. Fig. 1 shows the publication of GAN-related papers and their application in anomaly detection from January 2014 to September 2021.

This work focuses on the research progress and application of GAN in anomaly detection. At present, many studies are devoted to the summary of anomaly-detection techniques [11–13] and summarize the research of anomaly detection based on deep learning; however, they only briefly introduce the application of GAN for anomaly detection. Additionally, current research [14] has only summarized three main GAN-based anomaly-detection methods and their performance comparison on public datasets, but there is no global summary of the latest developments and future directions of GAN. Therefore, in this review, we will untangle the theoretical evolution of GAN, the development of anomaly-detection methods, and its implementation in specific applications in detail. The summary of this work will be of immense significance and will help researchers fully understand the latest developments, challenges, and future research directions of GAN in anomaly detection.

1.1. Rethinking anomaly

1.1.1. Concepts of anomaly

In detection tasks, the concept of an anomaly has several similar concepts, such as outliers [15], out of distributions (OOD) [16], novelty [17] and deviations [18]. Table 1 summarizes the different concepts and their corresponding definitions used in different research fields. We classify the concepts used in the literature into three categories: anomaly/anomalies, outlier/OOD data, and novelty. Most of these concepts have similar definitions: in general, anomalies are objects (e.g., observations, patterns, cases, and points) that are inconsistent with well-defined data (normal/in-distribution instances, or the majority of objects) [19,12,1,10,20–22]. For example, they do not conform to the clearly defined behavior [1], deviate from most observations [10], or are notably different from other data points without imitating them [23]. Previous researchers [24,13] have categorized the types of anomaly according to the type and intrinsic properties of the data. Chalapathy et al. [12] divided the types of anomaly into point, contextual or conditional, and collective or group with respect to data correlation. In recent research, based on whether the training data belong to the in-distribution, Bulusu et al. [11] subdivided anomalies into two types: unintentional and intentional. Based on the above, the conditions for judging anomalies can be summarized into two aspects: one assumption is that the normal samples have the same feature distribution in latent space, and the other is that the distribution of abnormal samples is too different from the normal data.

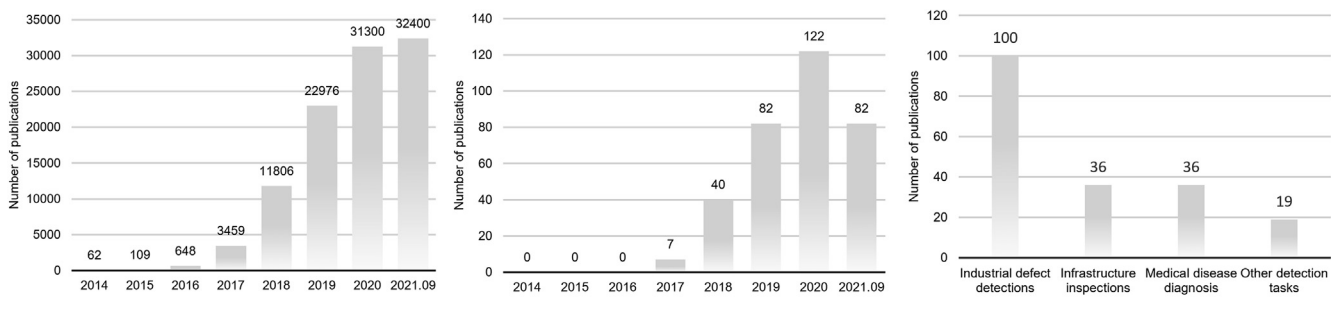
However, the definitions of these concepts remain ambiguous and conflicting. Previous research [25] recognized and analyzed that there are a few subtle differences between outliers and anomalies, and yet the authors still use them interchangeably in their review. Other researchers [26] holds that novelty is the data that does belong to the expected (i.e., normal) region in feature space, which is the same definition as for anomaly and outlier. However, even other researchers [1,17] consider that novelties are not necessarily anomalies, although the solutions and methods used in novelty detection, anomaly detection, and outlier detection are often similar. Moreover, researchers [27] have underlined the differences between anomaly, novelty outlier, and rare event-detection terms, and proposed a one-to-one assignment of them to learning scenarios under supervised classification. However, there is still no consensus on whether these related concepts can be extended to unsupervised learning, few-shot learning, transfer learning, and other scenarios.

1.1.2. Ideal distribution and model distribution

As can be seen from the above section, different scholars have different definitions of anomalies in different scenarios. To discuss anomalies in the context of unsupervised learning, this review advances a reconsideration of the definition of anomaly, as show in Fig. 2. One of our basic viewpoints is that the ideal distribution ***N*** of normal samples in the feature space and the model distribution ***M*** learned by the normal samples are two concepts that need to be distinguished. Here, we define the ideal distribution of normal samples as an objective entity; however, it cannot be fully recognized by humans. For example, people usually do not think of adding Tom (a character from the cartoon *Tom and Jerry*) when constructing the sample set of "cats", but few people would deny that Tom is indeed a cat. The model distribution is a feature distribution learned by limited human cognition (because the samples are collected based on this limited cognition). Thus, the model distribution is inevitably biased and the definitions of normal, novelty, and anomaly will vary depending on the distribution used.

In Fig. 2, we want all the normal samples to be constrained in an ideal compact distribution ***N*** (or a manifold) in feature latent space, as shown in the region surrounded by abnormal samples, and the normal samples and the abnormal samples can be significantly distinguished by the distribution boundary. The purpose of anomaly detection is to learn a model and map the normal samples to an ideal distribution to exclude abnormal samples. However, the distribution learned by the model is distorted in latent space due to:

1) Model distribution bias. The model distribution form is mismatched to the ideal distribution form (e.g., insufficient training data, inaccurate distribution constraints, the ideal distribution



(a) Publication of papers on GAN (b) GAN-based publications in anomaly detection (c) GAN-based publications in typical detection tasks

Fig. 1. Publication of papers on GAN and its application in anomaly detection from 2014.01 to 2021.09.

Table 1
Different research fields use different concepts and give corresponding definitions, some of which are ambiguous and conflicting. (CSUR: ACM Computing Surveys, EURASIP: European Association for Signal Processing)

Used concepts	Title	Venue	Definitions
Anomaly/ Anomalies	Anomaly Detection: A Survey [1]	*CSUR*	Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior (the same as [15]).
	An introduction to outlier analysis [19]	*Outlier Analysis, Springer*	Anomalies are also referred to as abnormalities, deviants, or outliers in the data mining and statistics literature.
	A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data [20]	*PloS ONE*	Anomalies are known to have two important characteristics: (1) Anomalies are different from the norm with respect to their features. (2) They are rare in a dataset compared to normal instances.
	A comparative evaluation of outlier detection algorithms experiments and analyses [28]	*PR*	Anomalies can be defined as observations which deviate sufficiently from most observations to consider that they were generated by a different generative process.
	Deep learning for anomaly detection: A survey [12]	*ArXiv*	Anomalies are also referred to as abnormalities, deviants, or outliers in the data mining and statistics literature (the same as [19]).
	Anomaly Detection Methods for Categorical Data: A Review [22]	*ACM Comput*	Anomalies are a minority of objects (observations, cases, or points) that are inconsistent with the pattern suggested by the majority of objects in the same dataset.
	Deep Learning for Anomaly Detection: A Review [13]	*CSUR*	Anomaly detection addresses minority, unpredictable/uncertain and rare events. The types of anomalies are explored: Point anomalies, are individual instances that are anomalous the majority of other individual instances. Conditional anomalies, a.k.a. contextual anomalies, also refer to individual anomalous instances but in a specific context. Group anomalies, a.k.a. collective anomalies, are a subset of data instances anomalous as a whole with respect to the other data instances.
	On the Nature and Types of Anomalies: A Review of Deviations in Data [24]	*ArXiv*	Using the intrinsic properties of the data to define and distinguish between the different sorts of anomalies, and presenting the four fundamental data-oriented dimensions employed to describe anomalies: data type, cardinality of relationship, data structure and data distribution.
	A Unifying Review of Deep and Shallow Anomaly Detection [21]	*Proceedings of the IEEE*	An anomaly is an observation that deviates considerably from some concept of normality, they may be termed unusual, irregular, atypical, inconsistent, unexpected, rare, erroneous, faulty, fraudulent, malicious, unnatural, or simply strange – depending on the situation.
	Anomalous Example Detection in Deep Learning: A Survey [11]	*IEEE Access*	In deep anomaly detection, anomalous samples are for testing, which do not conform to the distribution of the training data, they can classify into unintentional (novel and out-of-distribution examples) and intentional (adversarial examples).
Outlier/ OOD data	Outlier detection: A survey [15]	*CSUR*	Outliers, as defined earlier, are patterns in data that do not conform to a well-defined notion of normal behavior, or conform to a well-defined notion of outlying behavior.
	Outlier detection approaches for wireless sensor networks: A survey [18]	*Computer Networks*	In the context of wireless sensor networks, outlier also known as an anomaly or deviation is considered for identifying unusual behavior when compared to the majority of sensor readings.
	Progress in outlier detection techniques A survey [23]	*IEEE Access*	Outlier is generally considered a data point which is significantly different from other data points or which does not conform to the expected normal pattern of the phenomenon it represents.
	Deep Residual Flow for Out of Distribution Detection [16]	*CVPR*	The OOD detection problem seeks to assign a confidence score to the classifier predictions, such that classification of OOD data would be given a lower score than in-distribution data.
Novelty	Review of novelty detection methods [26]	*MIPRO*	Novelty (anomaly, outlier, exception) is a pattern in the data that does not conform to the expected behavior.
	A review of novelty detection [17]	*EURASIP*	Novelty detection can be defined as the task of recognizing that test data differ in some respect from the data that are available during training.
	Deep learning for anomaly detection: A survey [12]	*ArXiv*	Novelty detection is the identification of a novel (new) or unobserved patterns in the data (the same as [26]). The novelties detected are not considered as anomalous data points; instead, they are been applied to the regular data model. And novelty score may be assigned for these previously unseen data points, using a decision threshold score, the points which significantly deviate from this decision threshold may be considered as anomalies or outliers (the same as [17]).

may not exist in an inappropriate feature space, etc.). This bias of model distribution learning under biased cognition is inevitable because it is impossible for humans to completely identify the ideal distribution.

2) Model estimation error. Even if the model distribution form is correct, the model parameters may be inaccurate or not fully optimized (e.g., unbalanced training data, parameter initialization sensitivity, local maxima of the learning objective, etc.). This error can be reduced with the improvement of estimation methods; however, the process of making it converge to zero is a continuous effort.

Therefore, the model distribution will encompass some abnormal samples and exclude some normal samples. Furthermore, to reduce false negatives (i.e., abnormal samples that are regarded as positive samples and normal samples that are regarded as negative samples), the decision boundary used to detect anomalies is usually within the model distribution. The determination of anomalies is based on the model decision range $\boldsymbol{B}$ rather than the model distribution range $\boldsymbol{M}$($\boldsymbol{B} \subseteq \boldsymbol{M}$). Thus, given a sample p, there are five possible categories, as shown in Fig. 2:

- Normal samples that are within the decision boundary, such as $p_1 \in \boldsymbol{B}$

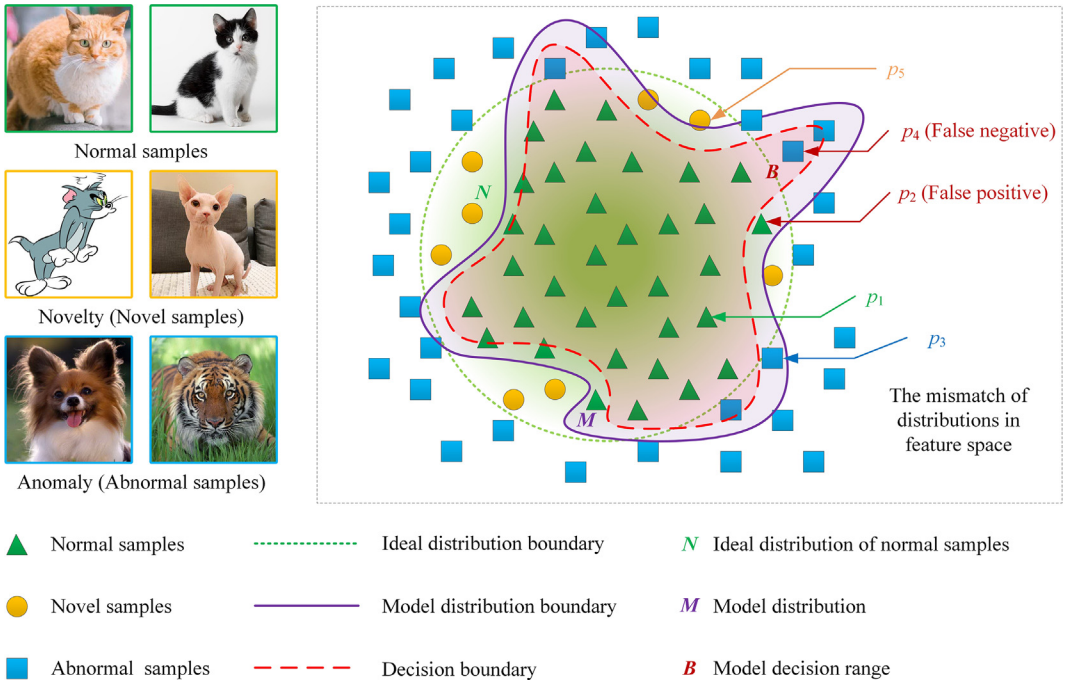


Fig. 2. Rethinking anomaly: The difficulty of anomaly detection comes from the mismatch of distributions in latent space due to model distribution biases and model estimation errors. Using the example of cats, the "cats" in the ideal distribution are the normal samples, and the "dog" and "tiger" outside the ideal distribution are distinguished as anomalies. However, there is a mismatch between the model distribution and the ideal distribution. Some abnormal samples are considered normal, some normal samples are considered abnormal, and the novelty "cats" outside the model distribution but within the ideal distribution may also be misjudged.

- Normal samples that are within the model distribution but out of the decision boundary (false positive), such as $p_2 \notin \boldsymbol{B}$ and $p_2 \in \boldsymbol{M}$
- Abnormal samples that are out of the ideal distribution and the decision boundary, such as $p_3 \notin \boldsymbol{B}$ and $p_3 \notin \boldsymbol{N}$
- Abnormal samples that are out of the ideal distribution but within the decision boundary (false negative), such as $p_4 \in \boldsymbol{B}$and $p_4 \notin \boldsymbol{N}$
- Novel samples that are within the ideal distribution but out of the model distribution, such as $p_5 \in \boldsymbol{N}$and $p_5 \notin \boldsymbol{B}$

The question that arises is, should the novel sample p_5 be seen as normal or abnormal? In other words, does the distribution in OOD refer to the ideal distribution or the model distribution? Alternatively, should the classification of p_5 be regarded as an error in Fig. 2?

These questions demonstrate that novelty cannot be simply regarded as an anomaly, and the definition of novelty and anomaly is related to the perspective adopted by researchers.

1.1.3. Human perspective and machine perspective

As Fig. 3 shows, we surveyed the definition of normal and abnormal from the human perspective and the machine perspective, respectively. From a human perspective, normal samples and abnormal samples should belong to two different sets, with one sample being either normal or abnormal (i.e., one hot labeled). In a normal sample set, even if there are novel samples that humans are not aware of, these novel samples should be naturally classified as normal. On the contrary, from the perspective of machine vision, normal and abnormal are two overlapping probability distributions after the model distribution is established. The decision boundary of the model can only distinguish all the samples outside the boundary as abnormal. Whether novel samples and abnormal samples can be distinguished depends on whether the distribution learned by the model can provide more discrimination information.

Now we can address the remaining questions left over from the previous section. Whether novel sample p_5 is considered abnormal depends on whether researchers employ a human perspective or a machine perspective. Using a human perspective, the ideal distribution is the D in OOD, novelty is also a normal outcome, and the classification error of p_5 is Type I Error (false positive) in Fig. 2. Using a machine perspective, the model distribution is the D in OOD, novelty is a kind of anomaly, and the classification of p_5 is not an error in Fig. 2.

We observe that when the model decision boundary can encompass all the normal samples, that is, when the model distribution can encompass the ideal distribution ($\boldsymbol{N} \subseteq \boldsymbol{B}$), human perspectives and machine perspectives can achieve the greatest agreement. At this point, the novel sample disappears, and the controversy over the novelty discrimination also disappears. However, there is no guarantee that novelty will not occur in practice due to model distribution biases and model estimation errors. Researchers must make a trade-off between the accuracy and generalization of the model distribution.

1.1.4. Criteria for discussing anomaly

To reiterate, in this review, we agree with the definitions of anomaly in most papers [19,18,11,12,15,1,20–23], but we object to simply equating anomaly with OOD data or novelty in practice. In this review, we define the anomaly as a departure from a pattern as identified by human cognition, resulting in perceptual changes (e.g., appearance damages, style changes, blur, or occlusion) or semantic changes (e.g., changes of category or reversals in behavior) to the pattern. Additionally, we think that when discussing an anomaly-detection task, the following criteria should be clarified first:

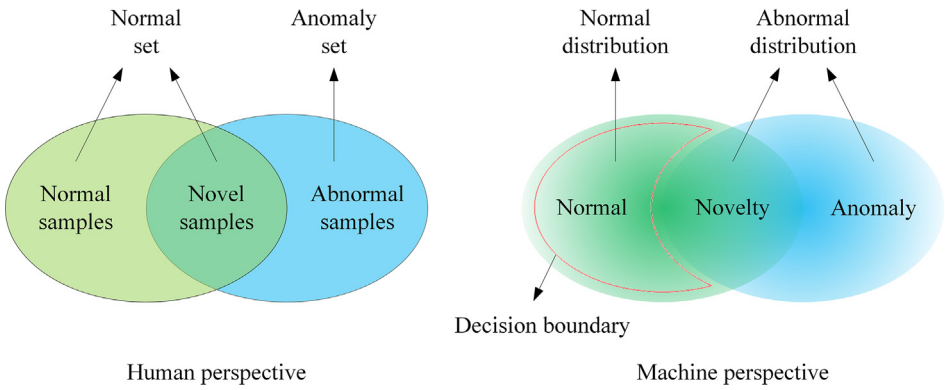


Fig. 3. Normal and anomaly are defined differently from the human and machine perspectives. Researchers should determine whether novelty should be classified as normal or abnormal based on its practical application.

- Whether the detection performance of the model is strongly correlated with training data or learning methods. Insufficient data or few-shot [29] learning are likely to result in model distribution biases that confuse many of the novel and abnormal samples. This is essentially a trade-off between accuracy and generalization. Therefore, researchers must distinguish the concepts of novelty and anomaly in research and discussion to better explore how to reduce the number of novel samples and improve the detection rate of abnormal samples.
- Whether the model is expected to have zero-shot learning [30] ability. Zero-shot learning requires the model to construct a hierarchical structure of the knowledge, and thus recognize new sample categories. In this case, novel samples and abnormal samples need to be studied and discussed under different representational metrics.
- Whether the model is expected to have a life-long learning [31] ability. Life-long learning requires the model to constantly learn new knowledge outside the model distribution, so it also needs to construct and constantly adjust different representations of novel and abnormal samples. The ability to generalize novel samples is an important performance indicator of life-long learning.

As an unsupervised learning method, the model distribution of GAN is more prone to model distribution biases and model estimation errors than other methods because it typically only learns from normal samples during anomaly-detection tasks. Therefore, GAN is more sensitive to the above criteria in different practical application scenarios. We believe that researchers should clarify the above criteria when studying GAN-based anomaly detection to avoid ambiguity and confusion, which is why we reconsider the topic of anomaly in this review. In response to the above reconsideration of anomaly, we now turn to a discussion of the challenges of anomaly-detection tasks.

1.2. Challenges

At present, DAD faces several key challenges. For example, it faces the problem of data imbalance among normal, abnormal, and novel samples. As mentioned in the previous section, an anomaly can be simply defined as a pattern that does not meet expectations. Therefore, the ideal method is to define an In-Distribution that represents normal samples and then determine any sample outside the distribution as an anomaly [15]. A previous review on the topic [13] provided an excellent summary of the main problems and complexities that lead to the ineffectiveness of deep learning in anomaly detection. However, as described in the previous section, the criteria used for anomaly-detection tasks are different in different scenarios. A single metric (such as accuracy or recall) cannot fully describe the advantages and disadvantages of different methods in different detection tasks. Therefore, we summarize the four general key challenges faced by the detection task to provide researchers with a greater appreciation of anomaly detection research, as shown in Fig. 4:

- **Data imbalance:** There are usually extensive negative samples (normal samples), absent, or few positive samples (abnormal samples), and abnormal samples are difficult to obtain in practical applications. Furthermore, data imbalance not only refers to the imbalance between the number of positive and negative samples but also includes imbalances of the internal distribution of normal samples. For example, if there is a lack of sphynx cats in the training data, a model trained to identify cats will likely recognize sphynx cats as anomalies. In other words, the goal is that the machine perspective based on model distribution can be as consistent as possible with the human perspective. Therefore, the anomaly-detection model not only needs to learn the appropriate representation from the training data, but also needs to learn an inductive bias [32] that is similar to the priori under human perspective, according to other reasonable constraints.
- **Decision boundary ambiguity:** Ideally, the decision boundary of the model should be equal to the ideal distribution boundary. However, the distortion of the model distribution not only leads to ambiguity between the normal, abnormal, and novel samples, but also leads to different degrees of ambiguity at different distribution boundaries. Therefore, the model not only needs to widen the distance between normal and abnormal samples as much as possible in its learning representation, but also needs to adaptively use the most appropriate decision boundary for detecting different samples. Addressing this challenge requires more accurate model estimations and better representation learning methods.
- **Abnormal metric:** One of the significant advantages of anomaly detection is to guide the corresponding disposal methods according to the peculiarities of the abnormal situation. However, most of the existing methods can only measure the degree of anomaly in a way that is uninterpretable. The interpretability of the model affects its credibility, which is very important for the practical application of anomaly detection. Interpretable anomaly detection results are more beneficial to downstream tasks such as anomaly classification, anomaly localization, and cause analysis. Therefore, anomaly-detection model must be able to properly cluster, cascade, and compute representations to measure and infer anomalies quantitatively and interpretably.

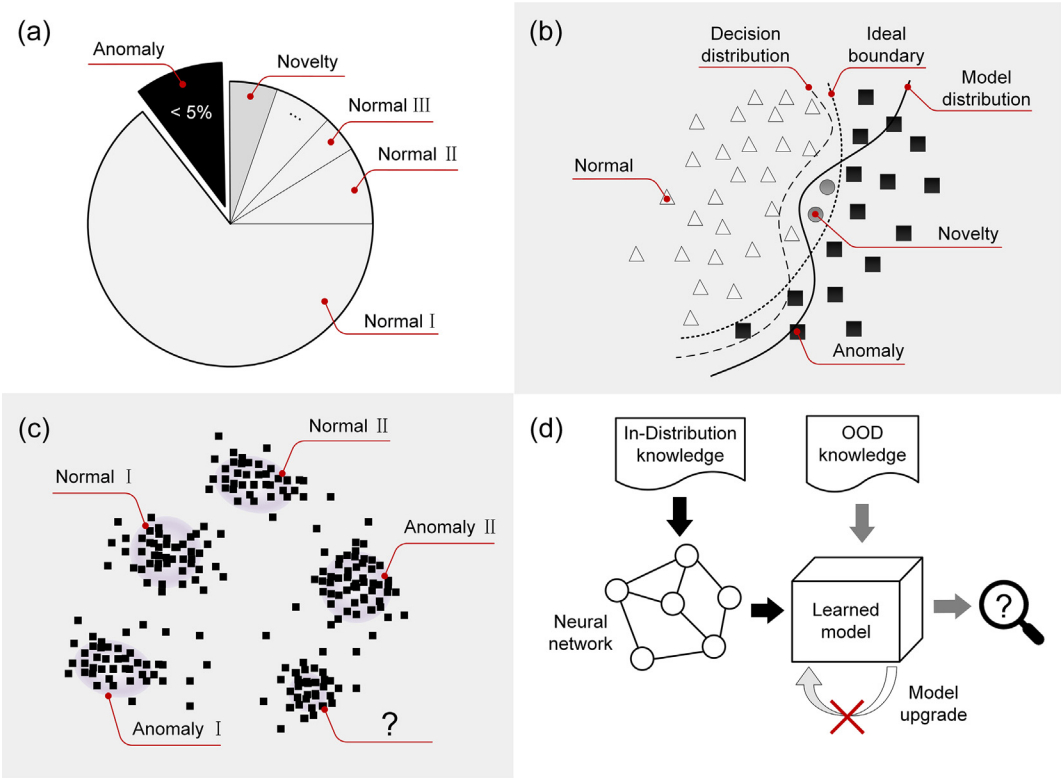


Fig. 4. Challenges of DAD: (a) Data imbalance; (b) Decision boundary ambiguity between normal, abnormal, and novel samples; (c) The un-interpretability of the abnormal metric; (d) The difficulty of OOD knowledge acquisition.

- **OOD knowledge acquisition:** The training of anomaly-detection model depends on a small number of abnormal samples or even only on normal samples, which makes it vulnerable to adversarial attacks of novel and unexpected abnormal samples. It is difficult for the model to learn the OOD knowledge of these adversarial samples during training. Nevertheless, any abnormal samples detected by the detection model during its operation should contain relevant OOD knowledge. Most existing methods ignore the acquisition of this OOD knowledge. Therefore, the anomaly-detection model must have the ability to incrementally learn, using life-long learning of OOD knowledge to upgrade the model and provide assistance for downstream tasks.

At present, anomaly-detection tasks based on GAN inevitably face the aforementioned challenges. In unsupervised, semi-supervised, and weakly supervised scenarios, GAN-based anomaly-detection methods face greater challenges. Although the latest data enhancement techniques significantly improve the representational learning ability of GAN [33,34], its potential for solving data imbalance problems remains unclear. The latest achievements of self-supervised learning offer hope for eliminating decision boundary ambiguity [35,36], but its advantages mainly pertain to the discriminative model rather than the generative model. The problem of the anomaly metric depends on the development of interpretable learning [37–40], which is a common difficulty shared across the field of deep learning. The problem of OOD knowledge acquisition can be best understood in the context of life-long learning and is still in its infancy [41]. This review attempts to disambiguate the existing relevant theories and technologies and discuss the corresponding research direction that should be taken.

1.3. Review organization and our contributions

In this work, we review the basic theory and the current technological applications of GAN-based anomaly detection. The review organization and our contributions are as follows:

- Section 1: Introduction. By distinguishing the ideal distribution from the model distribution, this section reconsiders the topic of anomaly from the human and machine perspectives, and provides three criteria for discussing the anomaly-detection task. Based on the above criteria, four key challenges of DAD are summarized: data imbalances, decision boundary ambiguity, anomaly metrics, and OOD knowledge acquisition.
- Section 2: Related works. After summarizing the development of DAD in recent years, the basic principles of GAN and its common variants and losses related to anomaly detection are introduced in this section. Additionally, this section reviews in detail the evolutionary history of the theory and applications of GAN-based anomaly detection.
- Section 3: GAN-based anomaly detection. This section discusses two basic theories of GAN-based anomaly detection in detail: Adversarial Learning and Inference and Self-supervised Learning of Representation. Then, the anomaly hypothesis and anomaly evaluation used in anomaly detection and location are introduced in detail.
- Section 4: Applications of GAN-based anomaly detection. This section summarizes the problems and applications of GAN-based anomaly detection in industrial defect detection, infrastructure inspection, medical diagnosis, and other key application areas, and provides a corresponding discussion.

- Section V: Discussion. This section discusses the current internal and external outstanding issues encountered by GAN-based anomaly detection and predicts and analyzes several future research directions.

This review provides a guide for understanding the principle, development, and application of GAN-based anomaly detection. Our goal is that, through this review, readers can understand the nature of the anomaly-detection problem and obtain the latest research developments in GAN-based anomaly detection. We hope that this review will provide a comprehensive reference for researchers and engineers who are eager to apply GANs for anomaly detection.

2. Related works

In this section, on the one hand, the research progress of DAD technologies from 2014 to October 2021 are summarized. Additionally, the development context and latest research direction of GAN-based anomaly detection with the improvement of GAN theory are described.

2.1. Research of deep anomaly detection

Deep learning, as a branch of machine learning, is designed to distinguish all kinds of samples by learning the data representation to obtain good performance and flexibility. Deep learning methods are frequently applied to anomaly detection, the purpose of which is to achieve anomaly detection through neural network learning of the feature representations or anomaly probabilities. The common DAD methods, such as those shown in Fig. 5, show better performance than traditional methods in solving challenging detection problems. Most DAD methods are based on the assumption that the testing data have the same distribution as the training data, and the goal of these methods is to improve the ability to judge whether the distribution of the former is different from the latter. A typical DAD method is designed to detect unseen data (that is, OOD or outlier detection). Based on the features of the training data, DAD methods are usually categorized into supervised, semi-supervised, and unsupervised approaches.

Supervised anomaly detection: Supervised methods must mark both normal and abnormal samples to learn the decision boundary from annotated instances, and the detector determines whether the test sample is normal or abnormal. Supervised methods are typically suitable for tasks wherein the abnormal features are easily labeled. In recent years, many studies have adopted different technologies to achieve supervised anomaly detection. Ma et al. [42] implemented anomaly detection using Local Intrinsic Dimensionality (LID) to characterize the dimensionality of adversarial samples. Lee et al. [43] proposed a simple yet effective method (Mahalanobis), which is applicable to any pre-trained Soft-Max neural classifier for detecting abnormal samples. Additionally, Jumutc et al.[44] found support for unknown high-dimensional distributions in the presence of labeling information for Supervised Novelty Detection (SND). Considering the robustness of Dropout Neural Network (Dropout-NN) to random disturbances, Feinman et al. [45] detected adversarial samples by examining the Bayesian uncertainty estimates, and by performing density estimations in the subspace of the deep features. To address the over-fitting problem that can be caused by supervised learning, Support Vector Data Description (SVDD) [46] was proposed to improve the generalization ability of one-class classification model.

Generally, compared with semi-supervised and unsupervised methods, supervised anomaly detection can achieve higher accuracy and faster detection speed. However, due to the need for extensive and accurately labeled training data, it is difficult to distinguish between normal and abnormal data with high-dimensional and complex features. Therefore, supervised methods are not as popular as semi-supervised and unsupervised methods.

Semi-supervised anomaly detection: For tasks with a large amount of training data and a high labeling cost, an effective method is to use semi-supervised learning to train a deep learning network by labeling only a part of the training data. In semi-supervised anomaly-detection tasks, the training data usually includes normal samples and partially labeled abnormal samples. Based on the Deep AE and *k*-Nearest Neighbor, Song et al. [47] propose a semi-supervised anomaly-detection model (DAE-KNN) for high-dimensional data. Wu et al. [48] proposed using semi-supervised deep Convolutional Recurrent Neural Networks (CRNN) for hyperspectral image classification for limited labeled data and abundant unlabeled data. Additionally, Perera et al. [49] designed a deep one-class classification (DOC) capable of learning the features of multi-labeled data for anomaly detection. The diversity of potential defects and the low availability of defective samples during manufacturing brings even more challenges. To address these challenges, a semi-supervised anomaly-detection method, dual prototype autoencoder (DPAE) [50], was proposed to distinguish anomalies on the industrial products surface because anomalies typically have different behavior patterns and there is generally much more unlabeled data than labeled data. Gao et al. [51] proposed a semi-supervised deep anomaly-detection method (ConNet) that utilizes large-scale unlabeled data and many labeled anomalies for modeling.

In general, semi-supervised learning can generally achieve better performance than other methods, even if there is little labeled data. However, because the features extracted from the latent layer

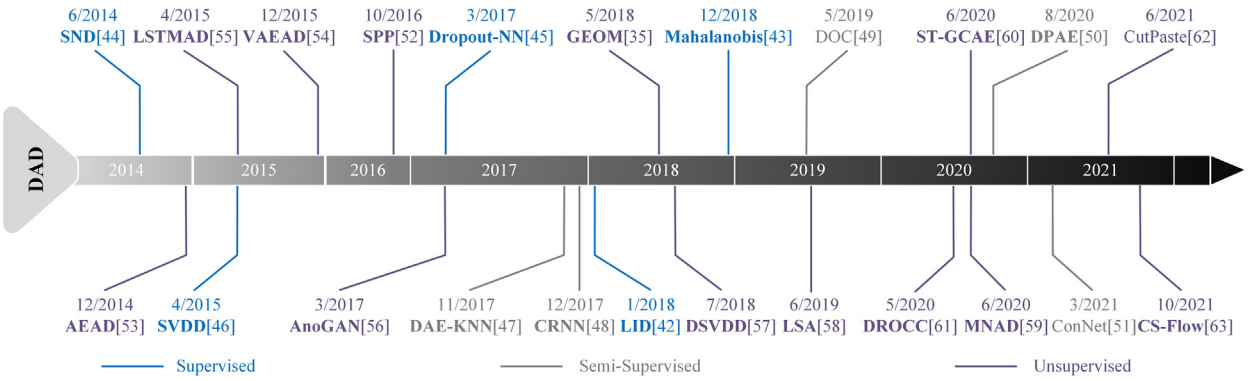


Fig. 5. Deep learning methods for anomaly detection from 2014 to October 2021.

may not represent abnormal instances, it is also susceptible to problems of over-fitting.

Unsupervised anomaly detection: Unsupervised learning is suitable for detection tasks wherein there are only normal samples, or it is difficult to obtain the label of abnormal data, and can distinguish normal samples from abnormal samples in latent space according to the feature distribution.

For unsupervised anomaly-detection tasks, Hendrycks & Gimpel [52] used confidence scores to determine if samples were OOD based on the Softmax Prediction Probability (SPP). AEAD [53] used autoencoders with nonlinear dimensionality reduction in the anomaly-detection task. To better learn the probability distribution and use it to calculate the anomaly score, VAEAD [54] utilized the reconstructed data generated by a variational autoencoder to analyze the causes of the abnormality, and verified that its performance was better than that of AE. Malhotra et al. [55] first applied stacked Long Short-Term Memory networks for Anomaly Detection (LSTMAD) in a time series. Schlegl et al. [56] proposed AnoGAN to learn a manifold of a normal image from latent space and subsequently recognize anomalies in new images. As the first deep one-class classification method for anomaly detection, Deep Support Vector Data Description (DSVDD) [57] learns neural network transformations with weights from the input space to the kernel space and attempts to map most of the normal representations into a hypersphere characterized by a predefined center and radius of minimum volume. Golan et al. [35] proposed Geometric Transformations (GEOM) for self-supervised one-class classification tasks. This method trains a multi-class neural classifier, which learns meaningful representation and salient geometrical features of normal images using different geometric transformations, to recognize anomalous images based on low SoftMax activation statistics. Another AE-based novelty-detection approach, Latent Space Autoregression (LSA) [58], fits an autoregressive model to the bottleneck layer by jointly training an AE paired with an additional autoregressive density estimator to learn the probability distribution of latent vectors for the normal samples by maximum likelihood principles.

In recent research, Park et al. [59] proposed using Memory-guided Normality for Anomaly Detection (MNAD) based on a CNN for video data, and Markovitz et al. [60] proposed a Spatio-Temporal Graph Auto-Encoder (ST-GCAE) to detect abnormal human postures. Goyal et al. [61] proposed a Deep Robust One-Class Classification (DROCC) for unsupervised anomaly detection. Their method assumes that the points from the normal class lie on a well-sampled and locally linear low-dimensional manifold, and functions by learning a representation to minimize the classification loss and then using a classifier to separate the normal samples from the anomalous samples. Aimed at constructing a high performance model for defect detection capable of detecting unknown anomalous patterns from an image without anomalous data, Li et al. [62] proposed a two-stage CNN for building anomaly detectors, which learns representations by classifying normal data from a data augmentation strategy (CutPaste). CS-Flow [63] proposed a novel fully convolutional cross-scale normalization flow that jointly processes multiple feature maps of different scales. This method preserves spatial arrangement, and as a result the latent space of the normalization flow is interpretable, which enables the method to localize defective regions in the image.

Unsupervised methods can learn commonalities in data from complex and high-dimensional space without requiring annotated training samples. However, they are sensitive to noise and abnormal data, and thus may be inferior in some contexts to supervised or semi-supervised methods. Therefore, methods to ensure the training stability of the network and obtain higher anomaly-detection accuracy will be the research focus for DAD methods in the future.

2.2. GAN for anomaly detection

2.2.1. The principle of GAN

In 2014, Goodfellow et al. [5] proposed GAN, and two-person zero-sum game theory is the core idea for their neural network training. The generation model skillfully uses an adversarial strategy to improve the effect of successive generations. The classic structure of the network is shown in Fig. 6.

A GAN consist of two parts: a generator and a discriminator. The generator attempts to capture the distribution of input data for the generated results, whereas the discriminator has two sources of input: the distribution of the generated data and of the real data. The optimization of GAN is a binary minimax adversarial problem. The purpose of optimization is that the generator attempts to generate results for which the discriminator finds it difficult to identify its source, and for the discriminator, the goal is to discriminate synthetic samples from real samples as accurately as possible. The output of the discriminator for real data is "1", and for generated data is "0". The loss function for guiding the discriminator training can be obtained as shown in Formula (1).

$$\max V(D,G) = \mathbf{E}_{\mathbf{x}\sim P_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (1)$$

where V denotes the output value of the loss function. G and D are the generator and discriminator of the GAN, respectively. $P_{\text{data}}(\mathbf{x})$ denotes the real data distribution, $P_{\mathbf{z}}(\mathbf{z})$ represents the distribution of generated data, and E is the mean.

For the generated data, the G attempts to be recognized by the D and the output is "1". Therefore, it can be concluded that the corresponding loss function of G is in the form shown in Formula (2).

$$\min V(D,G) = \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (2)$$

In general, Formulas (1) and (2) are combined as

$$\min_G \max_D V(D,G) = \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (3)$$

According to the theoretical analysis, in order to achieve the best generated quality, the generated data distribution of G should be consistent with the real data distribution. Therefore, a well-trained GAN should be applied to fit any sample distribution and determine whether a given new sample is in distribution.

2.2.2. Common variants and losses

The emergence of vanilla GAN was accompanied by problems of training instability (wherein the training of the discriminator diverges and the generator generates meaningless samples) and mode collapse (wherein the gradient of discriminator vanishes and the generator generates only samples with a few fixed modes) [5]. To allow the system to adapt to various tasks and compensate for the shortcomings of the vanilla GAN, various variants and losses have been proposed. Only the specific variations and losses that are related to anomaly detection are listed below.

DCGAN [64]. DCGAN provides a standard GAN architecture based on convolution. DCGAN verifies that the discriminator can be used for feature extraction in supervised learning tasks and that the generator can be used for semantic vector computation.

CGAN [65]. CGAN implements conditional generation by adding conditional information $\mathbf{y}$ to the generator and the discriminator. CGAN is a powerful tool for controlling the sample semantic because conditional information can be provided as labels, vectors, or even images. The loss function is

$$\min_G \max_D V(D,G) = \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x}|\mathbf{y})] + \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z}|\mathbf{y})))] \quad (4)$$

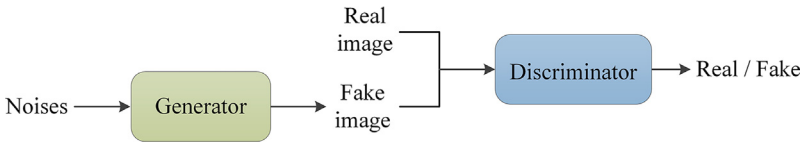


Fig. 6. Structure illustration of the generative adversarial network (GAN).

Info GAN [66]. Info GAN splits the input vector **z** into two parts: an interpretable hidden variable and an incompressible noise. By constraining the relation between them, the dimension of the hidden variable corresponds to the semantics of the generated sample. Info GAN strengthens control over the semantics of sample generation by introducing information theory, and its loss function is

$$\min_G \max_D V_I(D,G) = \min_G \max_D [V(D,G) - \lambda I(\mathbf{c}; G(\mathbf{z},\mathbf{c}))] \tag{5}$$

where **c** is the interpretable hidden variable, $I(\cdot)$ indicates the computation of mutual information, and λ is a hyper-parameter.

WGAN [67]. Previous analysis has demonstrated that the training of vanilla GAN is equivalent to minimizing the JS divergence [68] between the real distribution and the generated distribution. However, this divergence is not optimal. By minimizing the Wasserstein distance while satisfying the Lipschitz continuity [68], WGAN (Wasserstein GAN) theoretically solves GAN's problems of training instability and mode collapse. There are several variants of WGAN, the most common being WGAN-gp [8], whose loss function is

$$\begin{aligned}\min_G \max_D V(D,G) \quad &= \mathbf{E}_{\mathbf{z}\sim P_{\mathbf{z}}(\mathbf{z})}[D(G(\mathbf{z}))] - \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x})}[D(\mathbf{x})] \\ &+\lambda_{gp}\mathbf{E}_{\hat{\mathbf{x}}\sim P(\hat{\mathbf{x}})}\left[\left(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1\right)^2\right]\end{aligned} \tag{6}$$

where $\hat{\mathbf{x}}$ is sampled uniformly along a straight line between a pair of real and generated samples, and λ_{gp} is a hyper-parameter.

In most cases, researchers do not need to generate samples from noise but instead use an encoder (*En*) and decoder (*De*) instead of a generator to convert samples. Thus, the following losses were derived.

Reconstruction loss. This loss is often used in the training of an autoencoder for semantic maintenance [69] or conversion [70] at the pixel level, and its formula is

$$L_{\text{rec}} = \mathbf{E}_{\mathbf{x},\mathbf{y}\sim P_{\text{data}}(\mathbf{x},\mathbf{y})}[\|\mathbf{y} - De(En(\mathbf{x}))\|_n] \tag{7}$$

where $\|\cdot\|_n$ corresponds to the n-norm, and n is usually equal to 1 or 2. It is important to note that this loss takes many more forms. For example, it can be used not only for the reconstruction of samples, but also for the reconstruction of coding vectors.

Cycle/cyclic consistency loss. Cycle consistency loss [71], also known as cyclic consistency loss, guarantees that translated samples preserve the content of their input samples. Typically, the loss is based on a loop between a pair of generators G_1 and G_2 (or encoders and decoders). It is defined as

$$L_{\text{cyc}} = \mathbf{E}_{\mathbf{x}\sim P_{\text{data}}(\mathbf{x}),\mathbf{y}\sim P_{\text{data}}(\mathbf{y})}[\|\mathbf{x} - G_2(G_1(\mathbf{x}))\|_n + \|\mathbf{y} - G_1(G_2(\mathbf{y}))\|_n] \tag{8}$$

Similarly, its application is not limited to samples, but also hidden variables, masks, conditional information.

2.2.3. Evolution of GAN-based anomaly detection

GAN-based anomaly-detection methods that began with an intuitive assumption, evolved as the related theories evolved, incorporating new technologies and expanding their range of applications. Up to October 2021, the development of GAN-based anomaly detection can be roughly divided into three periods.

2016–2017: Early exploration of adversarial learning and inference

Ideally, after training is complete, the generator can generate normal samples, and the discriminator can identify abnormal samples. Based on this assumption, Schlegl et al. [56] proposed AnoGAN in 2017, and was the first to attempt to utilize GAN for anomaly detection. In training, only normal samples were used to learn the unsupervised manifold distribution in latent space. During inference, a loss function was defined to find the vector (Z) that was closest to the sample distribution in the manifold space through multiple back-propagation iterations. The generator output the reconstruction image using forward propagation and compared it with the original image to recognize the abnormal regions. Additionally, the output of the discriminator could be used as an abnormal value to determine whether the sample was an anomaly.

However, the framework of AnoGAN was based on assumptions rather than theoretical considerations. Moreover, AnoGAN must iteratively solve the code closest to the abnormal sample during inference, which cannot be directly calculated. Thus, can we not only fit a distribution, but also utilize the learned feature distribution to make inferences? In fact, as shown in Fig. 7, before AnoGAN, BiGAN [72] and ALI [73] both provided the theoretical basis for inference using GAN from two perspectives, and the core of adversarial feature learning and inference was to learn the feature distribution accurately. Additionally, CIAFL [74] provided the definition of the task faced by adversarial feature learning and inference: produce a data representation that maintains meaningful variations of the data while eliminating noisy signals. This definition precisely fits the requirements of anomaly detection. However, the corresponding theories have not played a direct role in guiding GAN-based anomaly detection, and there were only a few studies based on vanilla GAN during the same period [75–78].

2018–2019: Application of ALI

With the improvements observed in the theoretical work, researchers began to improve the GAN-based anomaly-detection methods based on the relevant theories of adversarial feature learning and inference. For example, Zenati et al. [79] proposed Efficient-GAN based on BiGAN and established a baseline model of anomaly detection based on GAN, which adds an encoder from image space to latent space on the basis of AnoGAN and considerably improves the inference speed of the network. GANomaly [80] further modifies the network structure and loss function to constrain the latent space. After ALICE [81] formally unified adversarial learning and conditional entropy as cyclic consistency, ALAD [82] utilizes its theory and used three discriminators to further limit the distribution of data in latent space and thereby improved the performance of anomaly detection.

The progress of anomaly detection based on GAN was accompanied by the technical development of GAN. Theoretical analysis showed that the limit of the vanilla GAN would lead to training instability and mode collapse. For this reason, some scholars studied methods of improving training stability in detection tasks, such as improving the stability of GAN in one-class multi-dimensional fault detection through AE and hyper-parameter selection [71] Since 2017, the training methods of GAN have progressed in two distinct directions. One is to use a regularization term for the gradient penalty in the loss function to make the discriminator meet the Lipschitz continuity, such as in WGAN [67]. The second is to

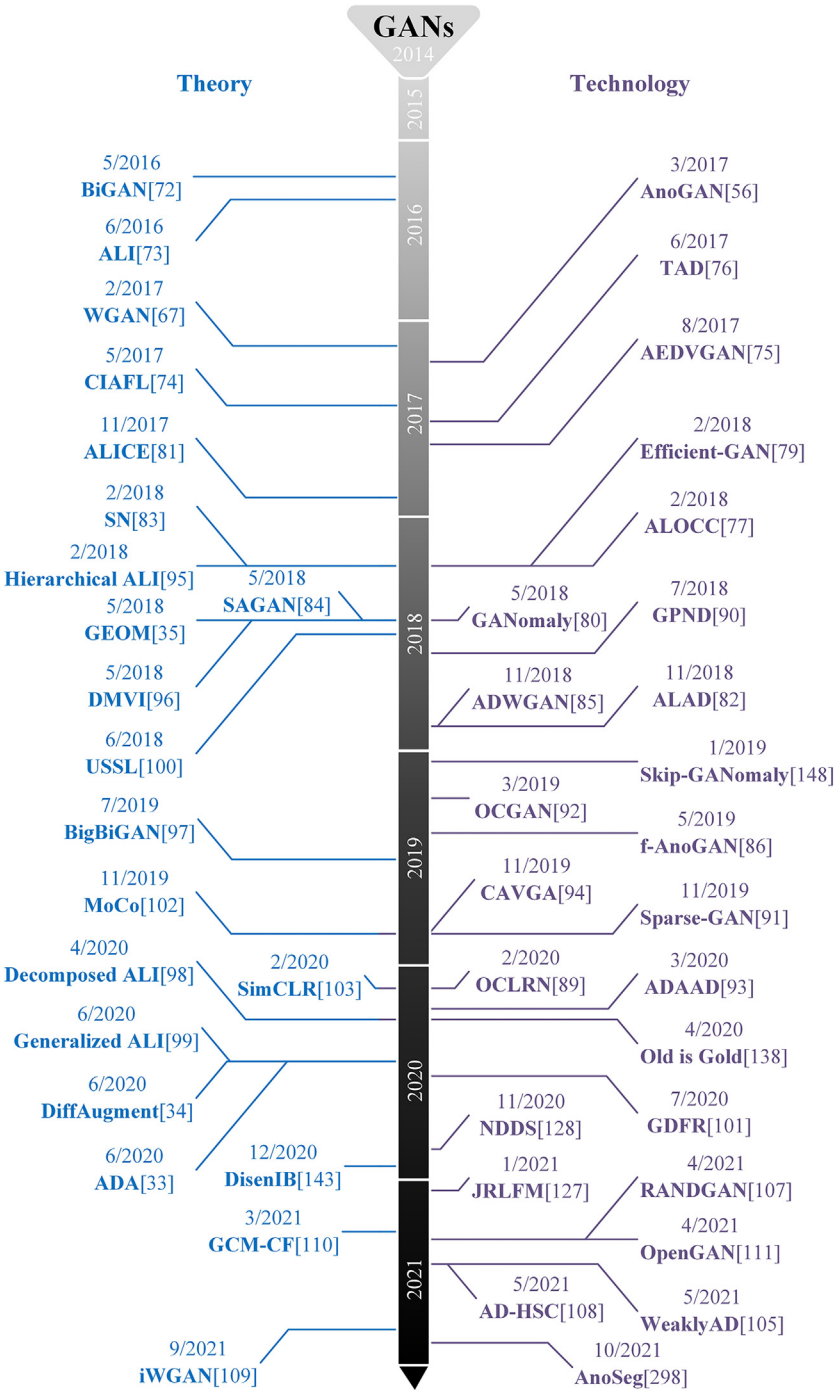


Fig. 7. Time series diagram of the theory and application of GAN-based anomaly detection. On the left of the timeline are the theoretical developments related to GAN-based anomaly detection, and on the right are important applications of GAN in the field of anomaly detection.

directly constrain the weights of the discriminator to meet the Lipschitz continuity, such as in spectral normalization [83]. Alternatively, since self-attention [84] has been put forward as a relevant technique, the development of attentional mechanisms began to lead the trend of research. These techniques considerably improved the stability of training and reduced mode collapse. GAN-based anomaly-detection methods immediately began to use the latest achievements of GAN to improve. For example, ADWGAN [85] and f-AnoGAN [86] learn the smooth representation of the variability of training data through WGAN, which improves the ability of anomaly detection. Other studies [87,88] utilized WGAN to train anomaly-detection networks to perform fault diagnosis and anomaly detection. To keep the model stable, OCLRN [89] utilized dual AE to restrict the space of latent representation in a discriminant manner. Since then, the research of GAN-based anomaly detection has been divided into two directions: one is to analyze the constraints of latent space deeply to better understand constraint methods from the perspective of manifold learning (GPND [90]), sparse regularization (Sparse-GAN [91]) and informative-negative mining (OCGAN [92]). The second approach is to explore more accurate methods of anomaly location, for example, ADAAD [93], which utilizes the class activation maps of

the discriminator to locate abnormal areas, and CAVGA [94], which proposes attention expansion loss to improve the accuracy of anomaly location.

However, the development of the theory of adversarial feature learning and inference stagnated during the period 2018–2019. During this period, only HALI [95] provided some new ideas in learning the distribution of latent space. However, DMVI [96] claimed that the VAE had its limitations, in that it cannot accurately learn the edge distributions of latent and visible space. Additionally, BigBiGAN [97] showed superiority in the field of representation learning based on BiGAN. This indirectly prompted researchers to use GAN for anomaly detection. Until 2020, the proposal of DALI [98] and GALI [99] showed that researchers had begun to re-investigate the theoretical research of adversarial feature learning and inference.

2018–2020: Rise of self-supervised learning

With increasing emphasis on unsupervised learning, researchers realized that the ability of representation learning directly affects the performance of anomaly detection. Additionally, related studies [100,33,34] have shown that data-enhancement tasks related to self-supervised learning can improve the representation learning ability of GAN. Hence, the related self-supervised learning was introduced into this field. GEOM [35] was a pioneering work that performed efficient representation learning through the prediction of geometric transformations and improves the performance of anomaly detection. Since then, the potential of self-supervised learning in the field of anomaly detection has become more apparent. However, the use of self-supervised methods in the anomaly detection domain focused on discriminative methods, most of which are only beneficial to the discriminator rather than generator of GAN. For example, GDFR [101] attempted to implement better representation learning using a generative method that assisted the discriminative method. Generative methods are generally worse than discriminative methods with respect to anomaly-detection performance because the discriminative pretext task is difficult to apply directly to the generator. Meanwhile, as a kind of discriminative self-supervised method, at the end of 2019, MoCo [102] and SimCLR [103] demonstrated that the contrastive learning method has excellent representational learning ability, which encouraged the emergence of anomaly-detection methods based on contrastive learning, such as CSI [104].

2021: More open field

With the popularization of new and advanced technologies, GAN-based anomaly detection is currently (as of October 2021) used in many fields. For example, WeaklyAD [105] proposes a weakly supervised discriminative learning with a spectral constrained GAN for hyperspectral anomaly detection (HAD). UI-GAN [106] proposes a novel GAN-based fall-detection method using a heart rate sensor and an accelerometer. RANDGAN [107] proposes a randomized generative adversarial network that detects the X-ray images of an unknown class (COVID-19) from known and labeled classes (Normal and Viral Pneumonia) without the need for labels and training data from COVID-19 X-ray images. Lastly, AD-HSC [108] uses a WGAN to detect anomalies in nearly one million optical galaxy images from the Hyper Suprime-Cam survey.

Recent theoretical studies on GAN-based inference, Zero-Shot Learning (ZSL), and Open-Set Recognition (OSR) have shown that GAN still has significant potential in representational learning and inference. For example, iWGAN [109] introduced a novel inferential Wasserstein GAN which is able to provide a measurement corresponding to the quality check of each individual sample. GCM-CF [110] presents a novel counterfactual framework by a counterfactual generation and inference model trained by WGAN. It provides a theoretical ground for balancing and improving the seen/unseen classification imbalances. OpenGAN [111] augments the available set of real open training examples with adversarial synthesized "fake" data and builds a discriminator over the features computed by the closed-world K-way networks. Extensive experiments have demonstrated that OpenGAN significantly outperforms prior open-set methods. With the help of the above research results, GAN may be able to further expand its role in anomaly detection.

3. GAN-based anomaly detection

This section summarizes the theoretical basis of GAN for anomaly detection and analyzes the applications of GAN in anomaly detection and the specific implementation for locating anomalies.

3.1. Theoretical basis

The successful applications of GAN-based anomaly detection depend on two theoretical bases: a learnable inference architecture and an effective representational learning ability. In early development, researchers focused more on the exploration of architecture. With improvements to the theoretical basis of GAN and the exhaustion of areas for architecture exploration, researchers have begun to turn to the study of representational learning ability and promote the introduction of related self-supervised learning methods.

3.1.1. Adversarial Learning and Inference

Although the relevant theoretical basis was not explicitly presented or quoted in AnoGAN, the theory of using GAN to detect anomalies was actually first identified in BiGAN [72] and ALI [73]. Essentially, BiGAN and ALI were different interpretations of the same set of theories. When applied to anomaly detection, ALI and BiGAN provided such clear theoretical support: once the sample distribution and latent variable distribution were correlated through strong constraints, the changes of latent variables could be used to predict whether the sample was abnormal.

For a vanilla GAN, the latent space captures semantic variation from the data distribution. Then the model can perform predictions according to the semantic latent representations. However, the vanilla GAN has no means of learning inverse mapping - projecting data back into the latent space. Both BiGAN and ALI attempted to implement this bidirectional mapping using a feature learning framework. Taking BiGAN and ALI as an example, as depicted in Fig. 8, in addition to the generator (G) from the vanilla GAN, BiGAN includes an encoder (E) which maps data x to latent representations $\mathbf{z}$. The discriminator (D) discriminates not only in the data space ($\mathbf{x}$ versus $G(\mathbf{z})$), but jointly in the data and latent space (tuples $(\mathbf{x}, E(\mathbf{x}))$ versus $(G(\mathbf{z}), \mathbf{z})$), wherein the latent component is either an encoder output $E(\mathbf{x})$ or a generator input $\mathbf{z}$. Additionally, the training objective is defined as

$$\min_{E,G} \max_{D} V(D, E, G), \tag{9}$$

where

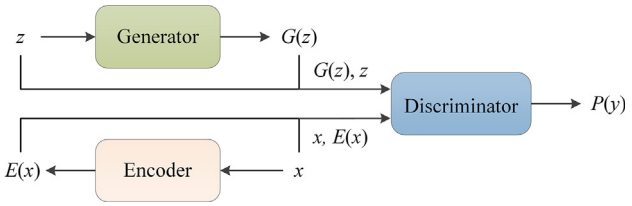


Fig. 8. Structure of Bidirectional GAN (BiGAN) and Adversarially Learned Inference (ALI).

$$V(D,E,G):$$
$$= \mathbf{E}_{\mathbf{x}\sim p(\mathbf{x})} \underbrace{\left[\mathbf{E}_{\mathbf{z}\sim pE(\cdot|\mathbf{x})}[\log D(\mathbf{x},\mathbf{z})]\right]}_{\log D(\mathbf{x},E(\mathbf{x}))} + \mathbf{E}_{\mathbf{z}\sim p(\mathbf{z})} \times \underbrace{\left[\mathbf{E}_{\mathbf{x}\sim p(G(\cdot|\mathbf{z}))}[\log(1-D(\mathbf{x},\mathbf{z}))]\right]}_{\log(1-D(G(\mathbf{z}),\mathbf{z}))} \tag{10}$$

The vanilla GAN only narrows the distance between the generated sample distribution and the real sample distribution. For the first time, BiGAN and ALI established a connection between the sample distribution and the latent space distribution, and provided an architecture for effective learning and inference in the latent space, from which GAN-based methods have a theoretical basis for anomaly detection.

Next, CIAFL [74], an invariant feature learning framework was introduced. It defined the tasks that adversarial feature learning faces: produce a data representation that maintains meaningful variations of data while eliminating noisy signals. This definition perfectly corresponds to the goal of anomaly detection because the anomaly can be seem as detrimental variations. Meanwhile, Efficient-GAN [79] investigated the use of BiGAN in anomaly-detection tasks and marked the time that the theoretical basis established by BiGAN and ALI officially began to support the development of anomaly-detection technology.

Further, although BiGAN is more widely known, adversarial learning inference has been more widely used for naming various variants of such learning frameworks because inference is the core function of feature representation and data reconstruction. The variants are as follows:

ALICE (ALI with Conditional Entropy) [81] ALICE investigates the non-identifiability issue of ALI, proposing to regularize ALI using the framework of conditional entropy (CE), and unified ALI/ BiGAN, CycleGAN [71]/ DiscoGAN [112]/ DualGAN [113] and Conditional GAN [65] as joint distribution matching. The objective of ALICE is the sum of ALI and the CE, regularized as shown in Formula (11):

$$\min_{E,G}\max_{D} V(D,E,G) = \mathscr{L}_{\text{ALI}}(\theta,\phi,\omega) + \mathscr{L}_{\text{CE}}(\theta,\phi) \tag{11}$$

where, θ,ϕ,ω are parameters of E,G, and D, respectively, $\mathscr{L}_{\text{ALI}}$ corresponds to the loss of ALI, and $\mathscr{L}_{\text{CE}}$ corresponds to the conditional entropy loss and can be replaced by the cycle-consistency loss, as its exact form depends on the task.

CE is approximated by additional discriminators in the ALICE because the cycle-consistency is an upper bound of the CE. ALICE stabilizes the learning of unsupervised bidirectional adversarial learning methods, but also increases the redundancy of the network.

HALI (Hierarchical ALI) [95]

HALI proposes a structure that uses multiple layers of Markov Kernels as the encoder, claiming that multiple layers of transformation performs better than a single layer. The hierarchical structure shown in Fig. 9 supports the learning of progressively more abstract representations as well as providing semantically meaningful reconstructions with different levels of fidelity. In Fig. 10, T are two teams of inverse feature transitions (encoders and decoders). The joint distribution of the encoder can be written as

$$q(\mathbf{x},\ldots,\mathbf{z}_L) = \prod_{l=2}^{L} q(\mathbf{z}_l|\mathbf{z}_{l-1}) q(\mathbf{z}_1|\mathbf{x}) q(\mathbf{x}), \tag{12}$$

whereas the joint distribution of the decoder is given by

$$p(\mathbf{x},\ldots,\mathbf{z}_L) = p(\mathbf{x}|\mathbf{z}_1) \prod_{l=2}^{L} p(\mathbf{z}_{l-1}|\mathbf{z}_l) p(\mathbf{z}_L). \tag{13}$$

The hierarchical structure of HALI reduces the reconstruction error of the generator and makes the network more sensitive to changes in the latent variables, which is conducive to the inference of sample changes.

DALI (Decomposed ALI) [98]

DALI explicitly matches prior and conditional distributions in both the data and code spaces, and puts a direct constraint on the dependency structure of the generative model. The DALI changes the original Encoder-Decoder structure to an $\mathbf{z}\to\tilde{\mathbf{x}}\to\tilde{\mathbf{z}}$ Decoder-Encoder structure. Unlike the original ALI, the *Discriminator* only discriminates between $G(\mathbf{z})$ and $\mathbf{x}$. DALI makes new assumptions about the distributions of *Generator* and *Encoder*:

- The *Generator* distribution: $p_\theta(\mathbf{z})p_\theta(\mathbf{x}|\mathbf{z})$; $\mathbf{z}\sim p_\theta(\mathbf{z})$, $\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{z})$.
- The *Encoder* distribution: $q_\phi(\mathbf{x})q_\phi(\mathbf{z}|\mathbf{x})$; $\mathbf{x}\sim q_\phi(\mathbf{x})$, $\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})$.

DALI decomposes the problem of minimizing $KL(p_\theta(\mathbf{x},\mathbf{z}),q_\phi(\mathbf{x},\mathbf{z}))$ into matching both the prior and conditional distributions, that is, to minimize

$$\mathbf{E}_{p_\theta(\mathbf{z})} KL(p_\theta(\mathbf{x}|\mathbf{z})\|q_\phi(\mathbf{x}|\mathbf{z})) + KL(p_\theta(\mathbf{z})\|q_\phi(\mathbf{z})) \tag{14}$$

In DALI, adversarial inference is incorporated into this framework and there is no parametric assumption on the conditional data distribution.The encoder of DALI only needs to output two vectors, $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$, that is, $E(\mathbf{x}) = (\mu(\mathbf{x}),\sigma^2(\mathbf{x}))$. Its final optimization problem is

$$\min_{E,G}\max_{D}\{V(D,G) + \lambda \mathbf{E}_{p_\theta(\mathbf{z})}[L(\mathbf{z},E(G(\mathbf{z})))]\} \tag{15}$$

where, λ is a hyper-parameter, DALI assumes $\mathbf{z}|\mathbf{x}\sim N(\mu(\mathbf{x}),\sigma^2(\mathbf{x}))$, and

$$\begin{aligned} & L(\mathbf{z},\mu(\mathbf{x}),\sigma^2(\mathbf{x})) := -\log q_\phi(\mathbf{z}|\mathbf{x}) \\ = & \ \tfrac{1}{2}\sum_{j=1}^{d}\left(\tfrac{(z_j-\mu_j(\mathbf{x}))^2}{\sigma_j^2(\mathbf{x})} + \log\sigma_j^2(\mathbf{x}) + \log(2\pi)\right) \end{aligned} \tag{16}$$

DALI outperforms ALI and ALICE and improves the generalization and inference ability of GAN.

GALI (Generalized ALI) [99]

GALI generalizes adversarial training approaches to incorporate multiple layers of feedback on reconstructions, self-supervision, and other forms of supervision based on prior or learned knowledge about the desired solutions, as shown in Fig. 10. GALI achieves this by constructing a multi-variable joint distribution and modifying the objective of the discriminator. By designing a non-saturating maximization objective for the generator-encoder pair, GALI proved that the resulting adversarial game corresponds to a

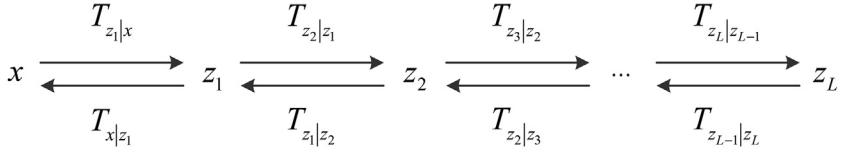


Fig. 9. The structure of HALI.

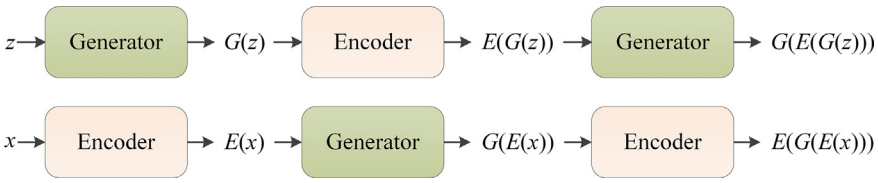


Fig. 10. The structure of GALI.

global optimum that simultaneously matches all the distributions. Additionally, GALI also introduced several techniques to provide self-supervised feedback for the model based on different properties, such as patch-level correspondence and the cycle consistency of reconstructions.

Simply put, GALI introduced multiple classification tasks into adversarial learning. The minimax objective with a multi-class classifier discriminator, following a straightforward generalization of ALI:

$$\min_{G,E} \max_{D} V(D,E,G) \tag{17}$$

where

$$\begin{aligned} V(D,E,G) := & \mathbf{E}_{\mathbf{x}\sim p_{\mathbf{x}}}[\log(D_1(\mathbf{x},E(\mathbf{x})))] \\ & + \mathbf{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(D_2(G(\mathbf{z}),\mathbf{z}))] \\ & + \mathbf{E}_{\mathbf{x}\sim p_{\mathbf{x}}}[\log(D_3(\mathbf{x},E(G(E(\mathbf{x})))))] \\ & + \mathbf{E}_{\mathbf{z}\sim p_{\mathbf{z}}}[\log(D_4(G(E(G(\mathbf{z}))),\mathbf{z}))] \end{aligned} \tag{18}$$

GALI also outperformed ALI and ALICE, and improved the generalization and inference ability of GAN. It also explored the utilization of self-supervised learning and pretrained models, which can enlighten future research.

3.1.2. Self-supervised learning of representation

Although GAN provides unsupervised learning and reasoning inference, its semantic representation capability is notably inferior to that of supervised learning. Accurate representation learning is beneficial for anomaly detection. First, accurate representation helps pull the normal samples closer and push the abnormal samples away. Second, reasonable representations can measure the degree of anomaly by similarity. In recent years, self-supervised learning has notably promoted the representation capability of networks in unsupervised environments, and thus that approach has been adopted by many anomaly-detection methods.

As shown in Fig. 11, contemporary self-supervised learning methods can roughly be broken down into two classes of methods in the field of computer vision: generative methods and discriminative methods. GAN can naturally be considered as a generative self-supervised learning method. The generator in generative methods learns representations by reconstructing x from the x_t transformed by $t(x)$, where $t(x)$ is the image transformation determined by the pretext task. On the other hand, the representations are learned by predicting the labels c_t provided by pretext task in discriminative methods. They may have better representation learning abilities for abstract semantics, and therefore they were introduced into the field of anomaly detection earlier.

Data enhancement and representation learning. The data-enhancement methods of generative methods and discriminative methods are different in self-supervised learning tasks. As a more traditional approach, generative methods focus on reconstruction error in the pixel space to learn representations, such as colorization [114], super-resolution [115], inpainting [116], and cross-channel prediction [117] and so on. Alternatively, discriminative self-supervised learning methods create (pseudo) labels using pretext tasks and learn representations through label predictions, such as geometric transformations [118], jigsaw puzzles [119], and context prediction [120].

Data enhancement is used to obtain self-supervised learning samples, but data enhancement alone can improve network performance. When considering only the role of data enhancement, there are at least three benefits of data enhancement in the training of GANs:

- Increasing the sampling density to make the learned distribution more consistent with the real distribution and thus improve the generalization ability.
- Learning invariant semantics from transformations to improve the effectiveness of the learned representations.
- Improving the robustness of the network in downstream tasks by enhancing specific capabilities of the network (e.g., denoising).

For anomaly detection, the most important function of data enhancement should be to facilitate the learning of semantic representation. Because GAN is divided into a generator and a discriminator, the effect of data enhancement on GAN needs to be analyzed separately. The effects of data enhancement and self-supervised learning on the performance of the discriminator were first studied in SS-GAN [121]; subsequently, DiffAugment (Differentiable Augmentation) [34] and ADA [33] investigated the impact of data enhancement on generator performance.

SS-GAN [121] exploits adversarial training and self-supervision, which can match equivalent conditional GAN on the task of image

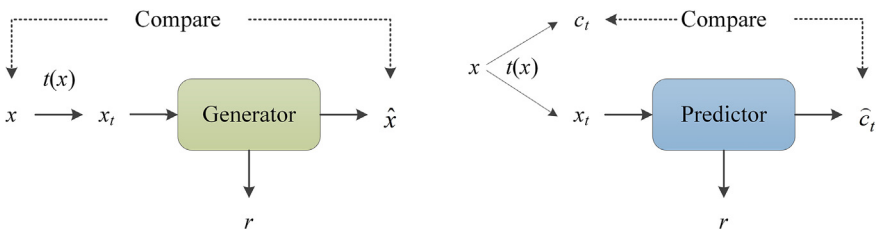


Fig. 11. Frameworks of generative methods and discriminative methods. (a) Generative methods learn representations r by the reconstruction of input x, (b) Discriminative methods learn representations r by the prediction of label c_t.

synthesis without having access to labeled data, as shown as Fig. 12(a). SS-GAN augments the discriminator with a rotation-based loss which results in the following loss functions:

$$L_G = -V(G,D) - \alpha \mathrm{E}_{x\sim P_G}\mathrm{E}_{r\sim R}[\log Q_D(\mathscr{R}=r|x^r)],$$
$$L_D = V(G,D) - \beta \mathrm{E}_{x\sim P_{data}}\mathrm{E}_{r\sim R}[\log Q_D(\mathscr{R}=r|x^r)]. \quad (19)$$

where $r \in \mathscr{R}$ is a transformation selected from a set of possible rotations. In this work we use $\mathscr{R} = \{0°, 90°, 180°, 270°\}$. Image x rotated by r degrees is denoted as x^r and $\log Q(R|x^r)$ is the discriminator's predictive distribution over the angles of rotation of the sample. $V(G,D)$ is the following value function for GAN training:

$$V(G,D) = \mathrm{E}_{x\sim P_{data}}[\log P_D(S=1|x)] + \mathrm{E}_{x\sim P_G(x)}[\log(1-P_D(S=0|x))] \quad (20)$$

where P_{data} is the true data distribution and P_G is the distribution induced by transforming a simple distribution $z \sim P_z$ using the deterministic mapping provided by the generator, $x = G(z)$, and P_D is the discriminator's Bernoulli distribution over the labels (real or fake).

In particular, the networks can collaborate on the task of representation learning, while being adversarial with respect to the classic GAN game. The role of self-supervision is to encourage the discriminator to learn meaningful feature representations which are not forgotten during training.

DiffAugment [34] adopts the differentiable augmentation for the generated samples, effectively stabilizing the training and leading to better convergence. It enables the gradients to be propagated through the augmentation back to the generator, regularizes the discriminator without manipulating the target distribution, and maintains the balance of training dynamics.

The processing of updating D and G is depicted in Fig. 12(b), for augmenting both real and fake samples and making generator G also focus on augmented samples. DiffAugment is defined as:

$$L_D = \mathbb{E}_{x\sim p_{data}(x)}[f_D(-D(T(x)))] + \mathbb{E}_{z\sim p(z)}[f_D(D(T(G(z))))],$$
$$L_G = \mathbb{E}_{z\sim p(z)}[f_G(-D(T(G(z))))]. \quad (21)$$

where the augmentation T must be differentiable.

Based on DiffAugment, ADA [33] applies an adaptive discriminator augmentation mechanism to significantly stabilize training with limited data. It does not require changes to loss functions or network architectures, and is applicable both when training from scratch and when fine-tuning an existing GAN. As shown in Fig. 12(c), it applies a diverse set of augmentations to every image shown to the discriminator, controlled by an augmentation probability $p \in [0,1]$, so that each transformation is applied with probability p or skipped with probability $1-p$.

Since the original studies, more research on data enhancement and GAN has been conducted [122–124]. These studies show that data enhancement can not only improve the discriminator's representational learning ability, but also improve the generator's representational learning ability through the designed training pipeline.

Discriminative self-supervised learning

Discriminative methods create (pseudo) labels using pretext tasks and learn representations by label predictions, such as image jigsaw puzzles [119], context predictions [120], and geometric transformation recognition [118].

As shown in Fig. 11(a), discriminative methods contain a *Predictor* that is trained to learn representations r by predicting the labels c_t from the x_t transformed by $t(x)$, where $t(x)$ is the image transformation determined by the pretext task. That requires the *Predictor* to try to predict the target label of input images by optimizing the loss between c_t and $\hat{c}_t$.

Most of these pretext tasks have been used for anomaly detection in recent years [125,126,35,100,101]. Although discriminative methods have not been widely combined with GAN, they have significantly deepened researchers' understanding of representational learning.

GEOM [35] successfully applied self-supervised learning to one-class classification and achieved state-of-the-art performance. It trains a multi-class model to discriminate between k-class geometric transformations ($\mathscr{T} = \{T_0, T_1, \ldots, T_{k-1}\}$, including horizontal flipping, translations, and rotations) applied on all given images. To identify which transformation is performed for each input, the classifier must learn salient geometrical features from normal images. Thus, this discriminative model effectively identifies anomalous images based on low SoftMax activation statistics during testing. USSL [100] demonstrated that self-supervision can benefit robustness in a variety of ways, including robustness to adversarial examples, label corruption, and common input corruptions. Additionally, self-supervision considerably benefits OOD detection for difficult, near-distribution outliers, so much so that it exceeds the performance of supervised methods. GFDR [101] aimed to determine if a given sample belonged to one of the classes used for training a model (i.e., the known classes). In Fig. 13(a), a generative model for all known classes is trained, and then the input is augmented with the representation obtained from the generative model to learn a classifier. The network aims to learn

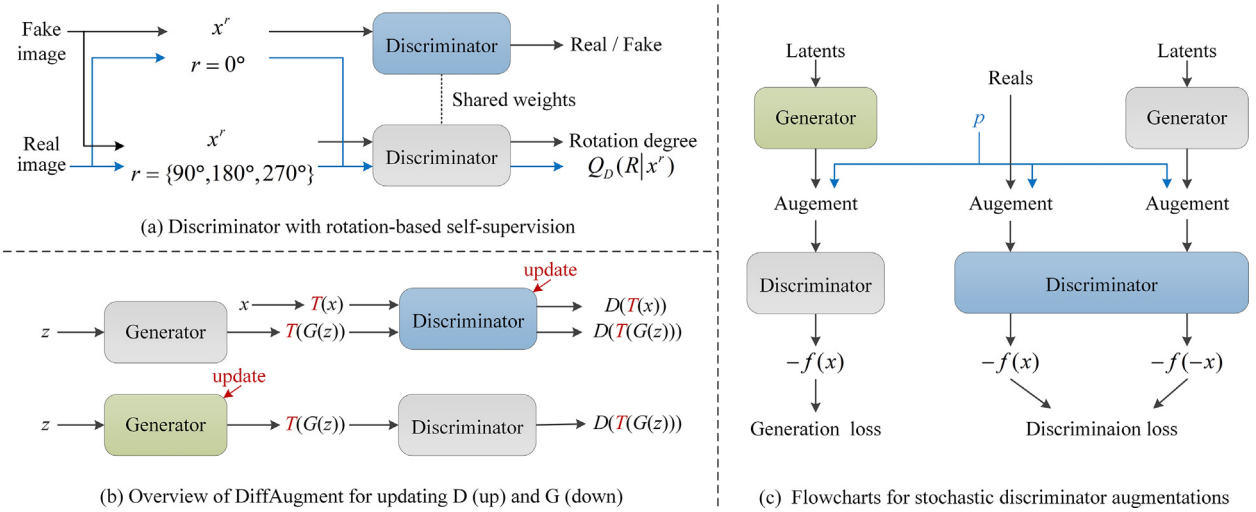


Fig. 12. Some data enhancement methods for GAN-based representation learning. a. SS-GAN [121], b. DiffAugment [34], c. ADA [33]. Gray indicates free.zing parameters.

to provide high classification probabilities when the image is from the correct class and when the input and the reconstructed image are consistent with each other. Second, self-supervision is used to force the network to learn more informative features when assigning class scores to improve the separation of classes from each other and from open-set samples.

JRLFM [127] is based on the combination of a generative framework and a one-class classification method, using the one-class data with a generative framework to learn features and augment the learned features with the corresponding reconstruction errors. Then, the augmented features are forced to take the form of suitable feature distribution that reduces the redundancy in the chosen classifier space using an adversarial framework as shown in Fig. 13(b).

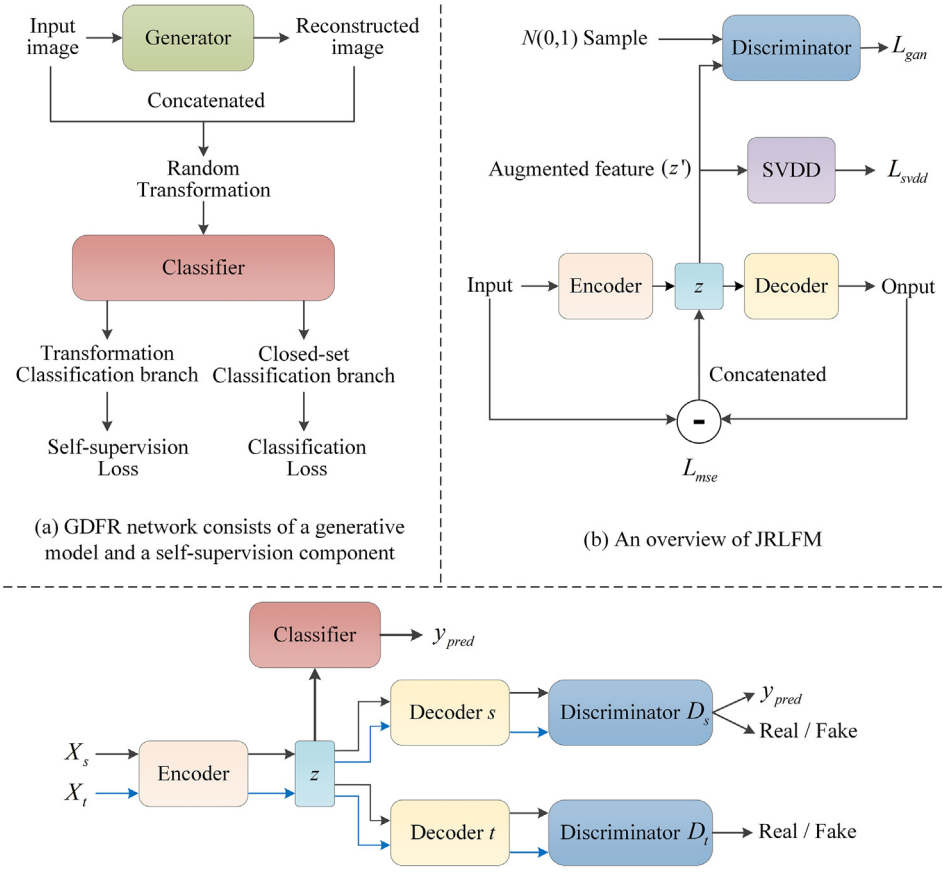
Additionally, NDDS [128] uses an integration of domain adversarial loss to improve generalization of multiple class novelty detection on the source domain(X_s) and target domain (X_t). The cross-domain decoders trained for novelty-detection task guide the shared feature encoder network to learn a common feature space, as shown in Fig. 13(c). The discriminator score of the reconstructed input is used for novelty detection at test time.

Generative self-supervised learning

Generative methods use a *Generator* to learn representations r by reconstructing x from the x_t provided by pretext task as shown in Fig. 11(b). This requires that the *Generator* learn the reconstruction ability of input images by computing pixel-level reconstruction loss between x and $\hat{x}$. As a more traditional approach, generative methods focus on reconstruction error in the pixel space to learn representations, such as colorization [114], super-resolution [115], inpainting [116], and cross-channel prediction [117]. However, using pixel-level losses can lead the model to overly focus on pixel-level details, rather than more abstract latent representations, thereby reducing their ability to model correlations or complex structure. As a kind of generative model, GAN training itself is a form of generative self-supervised learning. Most of the early GAN-based anomaly-detection methods only learn the representation through image reconstruction, such as AnoGAN [56], GANomaly [80] and ALAD [82]. Recently, researchers have recognized the limitations of image reconstruction and introduced richer pretext tasks to help generators learn representations, such as GAN+LBP (Local Binary Pattern) [129], Puzzle-AE [130] and ARnet [126].

GAN+LBP [129] is a novel defect-detection framework based on the training of positive samples. It establishes a reconstruction network, which can repair defect areas in the samples and then make a comparison between the input sample and the restored one to indicate the accurate defect areas. The model framework (in Fig. 14(a)) combines a GAN and an autoencoder for defect image reconstruction and uses LBP for image local contrast to detect defects.

Puzzle-AE [130] introduced an effective AE framework that combined adversarial robust training and self-supervised learning for anomaly detection. It is trained on solving puzzles of randomly permuted image patches and purposefully added FGSM [131] noises to images to alleviate the effect of low-level statistics such as edges, and each patch's channel-wise mean. The entire framework is shown in Fig. 14(b). This auxiliary task makes the AE net-

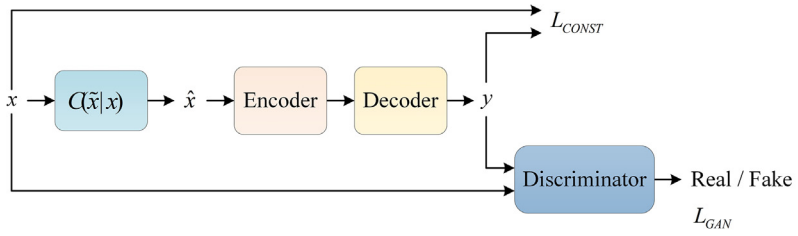


(a) GDFR network consists of a generative model and a self-supervision component

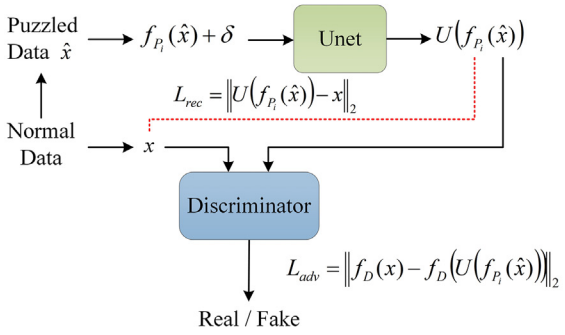
(b) An overview of JRLFM

(c) Illustration of NDDS which using cross-domain mappings for novelty detection

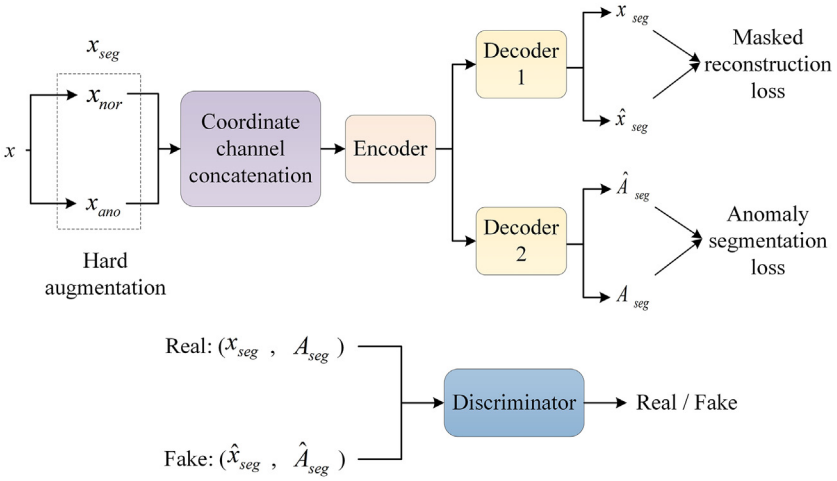
Fig. 13. Instances of discriminative self-supervised learning methods.



(a) The framework of GAN + LBP. It is an artificial defect module which can automatically generate a defective sample .



(b) Overview of Puzzle-AE framework. It passes puzzled input provided by pretext task and expects U-Net to reconstruct the right order images.



(c) Overview of the training process of the AnoSeg. AnoSeg generates reconstructed images and anomaly maps. To directly generate anomaly maps, AnoSeg applies three novel techniques: hard augmentation, adversarial learning, and coordinate channel concatenation.

Fig. 14. Network structures of generative self-supervised learning methods.

work focus on the essential properties of the normal class to solve the puzzle using adversarial training.

AnoSeg [132] proposes an anomaly segmentation network (AnoSeg) that can generate an accurate anomaly map as shown in Fig. 14(c). This method uses hard augmentation to change the normal sample distribution to generate synthetic anomaly images x_{ano} and reference masks A_{seg} for normal data. AnoSeg is trained in a self-supervised learning manner from the synthetic anomaly and normal data. Additionally, the coordinate channel, which represents the pixel location information, is concatenated to an input of AnoSeg to improve the performance of anomaly detection.

The performance gaps between GAN+LBP[129], Puzzle-AE [130], and AnoSeg [132] demonstrate that the representational learning ability of the generative self-supervised learning is extremely dependent on the design of pretext task. Research on pretext tasks suitable for anomaly detection remains an area for long-term exploration.

Contrastive self-supervised learning

As a type of discriminative method, contrastive learning methods treat each instance as a category, learning representations by contrasting positive and negative examples. Contrastive learning methods have recently shown considerable empirical success in computer vision tasks, starting with SimCLR [103], MoCo [102] and SimSiam [133]. Although contrastive learning is just emerging in the field of anomaly detection and has not yet been combined with GAN, its representational learning ability undoubtedly has immense potential [134,59,104].

Contrasting shifted instances (CSI) [104], specifically considers the SimCLR and proposes a new training method which contrasts distributionally shifted augmentations. $\tilde{x}_i^{(1)}$ and $\tilde{x}_i^{(2)}$ are two independent augmentations of x_i from a pre-defined family $\mathscr{T}$, namely, $\tilde{x}_i^{(1)} := T_1(x_i)$ and $\tilde{x}_i^{(2)} := T_2(x_i)$, where $T_1, T_2 \sim \mathscr{T}$. Then the SimCLR objective can be defined where each $\left(\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}\right)$ and $\left(\tilde{x}_i^{(2)}, \tilde{x}_i^{(1)}\right)$ are considered as query-key pairs while the others are considered negatives. Namely, for a given batch $\mathscr{B} := \{x_i\}_{i=1}^{B}$, the SimCLR objective is defined as follows:

$$\mathscr{L}_{\text{SimCLR}}(\mathscr{B}; \mathscr{T}) : = -\frac{1}{2B}\sum_{i=1}^{B} \mathscr{L}_{con}\left(\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \widetilde{\mathscr{B}}_{-i}\right) + \mathscr{L}_{con}\left(\tilde{x}_i^{(2)}, \tilde{x}_i^{(1)}, \widetilde{\mathscr{B}}_{-i}\right), \tag{22}$$

where $\widetilde{\mathscr{B}} := \left\{\tilde{x}_i^{(1)}\right\}_{i=1}^{B} \cup \left\{\tilde{x}_i^{(2)}\right\}_{i=1}^{B}$ and $\widetilde{\mathscr{B}}_{-i} := \left\{\tilde{x}_j^{(1)}\right\}_{j\neq i} \bigcup \left\{\tilde{x}_j^{(2)}\right\}_{j\neq i}$.

Additionally, augmentations S, which is referred to as *distribution-shifting* or *simply shifting transformations*, leads to better representation with respect to OOD detection when they are used as negatives in SimCLR.

CSI is defined by combining the two objectives: $\mathscr{L}_{CSI} = \mathscr{L}_{\text{con-SI}} + \lambda \cdot \mathscr{L}_{\text{cls-SI}}$.

Contrasting shifted instances (con-SI) loss: $\mathscr{L}_{\text{con-SI}} := \mathscr{L}_{\text{SimCLR}}\left(\bigcup_{S\in\mathscr{S}} \mathscr{B}_S; \mathscr{T}\right)$, where $\mathscr{B}_S := \{S(x_i)\}_{i=1}^{B}, \mathscr{S} := \{S_0 = I, S_1, \ldots, S_{K-1}\}$.

And classifying shifted instances (cls-SI) loss: $\mathscr{L}_{\text{cls-SI}} := \frac{1}{2B}\frac{1}{K}\sum_{S\in\mathscr{S}}\sum_{\tilde{x}S\in\tilde{\mathscr{B}}_S} - \log p_{\text{cls-SI}}(y^{\mathscr{S}} = S|\tilde{x}_S)$. where $\lambda$is a balancing hyper-parameter.

After the representation learned by the proposed training objective, a detection score is defined for detecting out-of-distribution: whether a given x is OOD or not. This score combines two features from SimCLR representations which are especially effective for detecting OOD samples: (a) the *cosine similarity* to the nearest training sample in $\{x_m\}$, i.e., $\max_m \text{sim}(z(x_m), z(x))$, and (b) the *norm* of the representation, i.e., $\|z(x)\|$. The detection score s_{con} for contrastive representation are:

$$s_{con}(x; \{x_m\}) := \max_m \text{sim}(z(x_m), z(x)) \cdot \|z(x)\|. \tag{23}$$

3.2. GAN for anomaly detection

With the development and optimization of the background theories, GAN has been gradually applied to anomaly-detection tasks in various areas after AnoGAN. At present, GAN-based image anomaly-detection methods typically train a network to learn the feature representation of positive (normal) samples. A previous review [13] summarizes generic normality feature representation learning based on GAN (as shown in Fig. 15): first, the training data $\mathscr{X}$ is input into the network to learn and obtain a feature extractor $\phi(\cdot)$, and the loss function based on the reconstruction or an anomaly measurement is used to detect abnormal samples. All samples are drawn on latent space learned from the normal space, and when they are used to generate images, the model forces the latent representation to reconstruct the like-normal images. Only the normal images can be reconstructed well and anomalies experience a high loss.

For GAN-based image-level anomaly-detection tasks, the common method is to input only non-abnormal samples to train the model and learn its feature distribution in latent space, and finally identify anomalies by comparing the differences between the reconstructed and abnormal samples by the generator, the discriminator, or both. It should be noted that the probability output by the discriminator can also be used as an anomaly score. However, the reliability of the discriminator for anomaly detection is controversial because it may degenerate after training [5].

3.2.1. Anomaly hypothesis

Before performing the anomaly-detection task, researchers need to determine the anomaly hypothesis. That is, training and test data are set according to the normal and abnormal samples distinguished by the task needs. For the current anomaly-detection methods based on GAN, the datasets are used differently for training and testing. For one-class anomaly-detection tasks [89,93,135–137,92,90,77,130,94,138,82,139], one class of datasets is usually trained as normal data to input to the network for training, and other classes are used for testing the accuracy of the network. For example, taking the digit "1" of MNIST dataset as normal, the rest of digits are regarded as abnormal samples.

However, in multi-class anomaly-detection tasks [80,79], one class is regarded as anomalous and the others as normal, that is, normal samples include multiple categories in the same datasets (like MNIST or Cifar-10), not only one category. Then, anomalies are detected as instances drawn from one class by training the model on normal classes. In novelty detection [128,92,90,140,104], one or more classes are used as the known categories and the remaining classes are regarded as novel categories. Only the known categories are used during training and the novel categories are used only to evaluate the models.

In the aforementioned GAN-based research, the training and testing data were derived from the same dataset, that is, normal and anomaly belonged to the same dataset. However, for out-of-distribution detection [141–147], multiple datasets are used for model training and testing. One dataset is used for training and testing as the in-distribution and the other datasets (one or more) are regarded as out-of-distribution for testing, and there is a large distribution gap between the two.

Therefore, the purpose of anomaly-detection tasks differs, and the allocation strategy of datasets correspondingly differs. Table 2 reviews the data-allocation techniques of the anomaly hypotheses described in the above methods.

3.2.2. Anomaly evaluation

Most GAN-anomaly algorithms are based on the generative model trained to learn the distribution of normal images in training and detect anomalies by analyzing reconstruction errors from the test data as an anomaly score [93]. Based on DCGAN [64], AnoGAN [56] used normal data to train the model and calculate errors from the trained generator and the discriminator as anomaly scores. GANomaly [80] used a larger distance metric from the learned data distribution in a constrained latent space to infer as the indicator of the abnormality of a given image. Similarly, Skip-GANomaly [148] adopted the anomaly score, combined with the reconstruction and latent representation score; a higher reconstruction metric was indicative of a deviation from normal distribution, that is, an anomaly.

To improve the learning ability of feature distribution, researchers have added an encoder to BiGAN [72]. Additionally, Efficient-GAN [79] defined a score function to evaluate if the reconstructed data has similar features in the discriminator as the real sample. Based on the adversarial learned features using three discriminators, ALAD [82] used reconstruction errors between a real sample pair and their reconstruction to determine if a data sample is anomalous.

After optimizing the training method of GAN to make the discriminator meet the limitations of Lipschitz, the performance of GAN-based anomaly detection is improved. AE [4] has been used as the learner of feature distribution in GAN and combines with the strong feature extraction ability of CNNs. ALOCC [77] applied

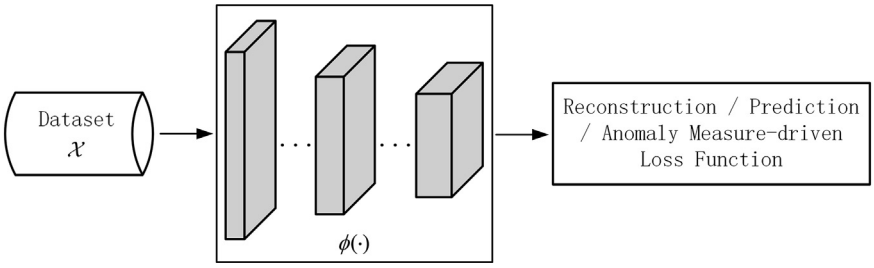


Fig. 15. Generic normality feature representation learning based on GAN.

Table 2
Data allocation for different anomaly hypotheses.

Data set		Tasks	Contents	References
Single data set	One-class	Anomaly detection	One class of the dataset is regarded as normal, the rest classes are abnormal.	CAVGA [94], ALOCC [77], ALAD [82], OCLRN [89], GPND [90], OCGAN [92], ADAAD [93], Puzzle-AE [130], Old is gold [138,135–137,139]
	Multi-class	Anomaly detection	One class of the dataset is considered as normal, and others are normal classes.	Efficient-GAN [79], GANomaly [80]
		Novelty detection	One or more categories are used as known classes and the remaining are considered as novel classes. Only the known classes are used during training and novel classes are used only for evaluation.	NDDS [128], CSI [104], OCGAN [92], DOCC [140], GPND [90]
Multiple data sets	/	Out-of-distribution detection	One dataset (all classes) is for network training, at testing time, examples from this dataset are viewed as in-distribution, and the data from other datasets as out-of-distribution.	[141–147]

adversarial learning to one-class novelty detection and detects novel samples by enhancing the inlier samples and distorting the outliers. Old-Is-Gold [138] also focused on adversarial learning to classify normal and abnormal data by training the discriminator to distinguish between good and bad quality reconstructions. In addition to image reconstruction residuals, f-AnoGAN [86] detected anomaly samples and (via an added residual on the discriminator features) yielded a reliable label for anomalies, as well as high anomaly scores for anomalous images and low scores for healthy images. Additionally, to obtain a clearer generated image, VAE [7] was used as a generator to learn the feature distribution of the input data. ADAAD [93] proposed a self-supervised masking method that specifically focuses on the detailed parts of images for anomaly detection; further, it designed the discriminator as a CNN whose output is a probability distribution, signifying the likelihood that the test image is from the normal class. CAVGA [94] computed the pixel-wise differences between input images and reconstructed images as the anomalous score, and empirically set a threshold to detect an image as anomalous.

In addition, GPND [90] focused on a novelty-detection method based on manifold learning, which captured the underlying structure of the inlier distribution and detects whether test samples are outliers by evaluating the probability distribution. OCGAN [92] was trained using only in-class samples and forced all out-of-class examples to generate realistic-looking examples and produce high Mean Squared Error (MSE). In the evaluation measures of OCLRN [89], the image reconstruction loss between the realistic image and the generated image was regarded as an anomaly score, and the distribution of anomalous samples can be separated from that of normal samples. Sparse-GAN [91] determined a threshold anomaly score for disease diagnosis from OCT images in the same manner as EAL-GAN [149]. Based on DCGAN, Wang et al. [150] proposed an unsupervised defect detection and location method (LDA) for texture images. A similarity analysis can be conducted between the reconstructed and original images, through which the probability of defect occurrence can be acquired. Moreover, the resulting residual maps are used to detect defects.

Table 3 shows the forms of the above methods to measure anomaly-detection performance and the related indicators. The GAN-based anomaly-detection methods mentioned above reconstruct abnormal data from normal data and the anomaly metric performance of their proposed methods have been evaluated on multiple public datasets. The $F1-score$ and the area under the curve (AUC) of the receiver operating characteristics (ROC) [151] are typically used as performance metrics for anomaly-detection tasks. Among them, $F1-score$ is defined as:

$$precision = \frac{TP}{TP + FP} \tag{24}$$

$$recall = \frac{TP}{TP + FN} \tag{25}$$

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}) \tag{26}$$

where TP, FN, FP and TN are the numbers of true positive, false negative, false positive, and true negative samples, respectively. Additionally, the ROC curve is plotted with true positive rates (TPR) against the false positive rates (FPR), the TPR and FPR are:

$$TPR = TP/(TP + FN) \tag{27}$$

$$FPR = FP/(FP + TN) \tag{28}$$

ROC is a probability curve and the AUC value represents the degree or measure of separability. AUC is calculated [152] as:

$$AUC = \frac{S_0 - M(N+1)/2}{M * N} \tag{29}$$

where M and N are the numbers of positive and negative examples, respectively, and $S_0 = \sum r_i$, where r_i is the rank of i_{th} positive example in the ranked score list [152]. An excellent model has an AUC close to 1, which indicates that it has a good measure of separability.

Table 3
Form of GAN-based methods for the evaluation of anomalies.

No.	Authors	Methods	Anomaly score	Indicator
1	Schlegl et. al. (2017.03)	AnoGAN [56]	Residual error + Discrimination error	AUC
2	H Zenati et al. (2018.02)	Efficient-GAN [79]	Reconstruction error + Discriminator error	F1-score
3	M Sabokrou et al. (2018.02)	ALOCC [77]	Reconstruction error	F1-score, AUC
4	S Pidhorskyi et al. (2018.07)	GPND [90]	Reconstruction error	F1-score, AUC
5	S Akcay et al. (2018.05)	GANomaly [80]	Reconstruction error	AUC
6	H Zenati et al. (2018.12)	ALAD [82]	Reconstruction error	F1-score, AUC
7	S Akçay et al. (2019.01)	Skip-GANomaly [148]	Reconstruction score + Latent representation score	AUC
8	P Perera et al. (2019.03)	OCGAN [92]	Reconstruction error	AUC
9	T Schlegl et al. (2019.05)	f-AnoGAN [86]	Reconstruction error + Discriminator error	F1- score, AUC
10	K Zhou et al. (2019.11)	Sparse-GAN [91]	Reconstruction error	AUC
11	S Venkataramanan et al. (2019.11)	CAVGA [94]	Reconstruction error	AUC
12	C Chen et al. (2020.02)	OCLRN [89]	Reconstruction error	AUC
13	Kimura et al. (2020.03)	ADAAD [93]	Reconstruction error + Discriminator error	AUC
14	MZ Zaheer et al. (2020.04)	Old-Is-Gold [138]	Discriminator error	F1-score
15	Chen et al. (2021.04)	EAL-GAN [149]	Discriminator error	AUC
16	Wang et al. (2021.11)	LDA [150]	Residual loss	F1-score

The above methods obtain better anomaly-detection performance and faster detection speeds by modifying the structure of networks, optimizing loss functions, and implementing data enhancement. For network training and testing, there are different data allocation in special tasks, which will be reviewed in Section 3.2.1.

3.3. GAN for anomaly location

Anomaly localization is a vital problem in visual detection and involves localizing anomalous regions. In GAN-based anomaly-detection tasks, it is generally not sufficient to simply judge whether there is an anomaly in a given input image; the model must also accurately identify the position of the anomaly. This is important for detecting, for example, whether the defect appears on an important work plane in industrial detection to consider whether it can be tolerated. It is also necessary to identify the location of the distress in large-area anomaly detection, such as with infrastructure, and it is necessary to identify the lesion region in medical image analysis.

Usually, ROC-AUC [153,151] and PRO-AUC [154,155] are commonly used as evaluation metrics in anomaly location and segmentation. The ROC-AUC assesses the best potential segmentation result with respect to normal and anomalous pixels, i.e., per pixel overlapping performance. The PRO-AUC metric attempts to measure the best possible segmentation performance across normal and anomalous regions at the region level, i.e., per region overlapping performance. In particular, the PRO-AUC measures a model's ability to segment out all possible anomalous regions equally, no matter what the size of a particular abnormal region is, whereas ROC-AUC may fail to measure this property because a sufficiently large, correctly segmented region can compensate for many wrongly segmented minor ones [156].

At present, GAN can generate realistic images, but anomaly-detection tasks also require the network to learn more abstract information from the training data and achieve anomaly location on a pixel-wise or image-level basis. Therefore, GAN-based anomaly-detection methods need to learn the details that appear consistently in normal classes and use those details to predict the anomalies. Usually, the methods of segmentation-based and attention-based anomaly location are adopted.

The results of anomaly location are usually visualized to improve the interpretability of the proposed model. For example, Sparse-GAN [91] proposed an Anomaly Activation Map (AAM) to visualize lesions in an anomaly-detection framework, and functioned by multiplying the latent feature map by an anomaly vector in a channel-wise fashion to obtain the anomaly activation map. Additionally, CBiGAN [157] addressed one-class anomaly detection of images using an improved BiGAN model with consistency regularization and visualized the pixel-wise absolute differences between input and restructured images. Fully Convolutional Data Description (FCDD) [38] presented an explainable deep one-class classification method, which used an explanation anomaly heatmap to show the performance of anomaly detection. From the results, the visualization of the abnormal region also corresponded to the quantitative result of the pixel-level location of the anomaly.

3.3.1. Segmentation-based anomaly location

At present, GAN-based anomaly location methods for image segmentation often perform unsupervised training on normal data without annotations. Segmentation-based anomaly location usually just compares the difference between the generated and the input image to identify anomalies and obtain better anomaly segmentation. For example, after unsupervised manifold learning of normal anatomical variability, AnoGAN [56] used the residual difference between reconstructed images and input images to identify anomalous regions within an image. Similarly, f-AnoGAN [86] also showed pixel-level anomaly segmentation results that were compared against the residual image. Healthy query images result in small deviations, whereas anomalous images were mapped to "reconstructions" yielding large deviations. f-AnoGAN yielded better pixel-level anomaly localization performance than other models, and its results can be further assessed by clinical experts as candidates for disease or through a subsequent classification approach.

3.3.2. Attention-based anomaly location

It is impossible to obtain accurate anomaly location results using exclusively residual images. Therefore, it is necessary to enhance the ability of GANs to learn features so that they can model important details. With the introduction of attention mechanisms and the incorporation of a self-attention mechanism into GAN to generate details based on feature location (SAGAN [84]), the anomaly-locating ability of GAN has been improved. In the latest research, attention mechanisms have been gradually introduced into GAN to capture geometric or structural patterns that occur consistently in some classes to identify and locate anomalies.

The attention mechanism in machine learning is a simulation of human observation mechanism for external things. When humans observe external things or objects, they generally do not see the thing as a whole, instead tending to selectively acquire some important parts of the observed object according to their needs. Attention can then combine information from different areas to form an overall impression of what is being observed. Therefore,

an attention mechanism can help the model assign different weights to each part of the input, thereby allowing it to extract more critical and important information and help the model make more accurate judgments.

For example, ADAAD [93] added an adversarial discriminative attention to GAN, allowing it to perform robust anomaly detection without prior anomaly samples. Additionally, the Class Activation Map (CAM) [158] has a discriminator which focuses on the local discriminative attributes to improve the robustness of network against background noise.

CAVGA [94] introduced attention expansion loss, such that the feature representation of the latent variables encode all the normal regions, and the network encourages the attention map generated from the latent space to cover the entire normal image. Then, the areas where the attention map does not focus are the abnormal regions from the input image.

Therefore, the introduction of attention into GAN can benefit many applications, such as by improving the contrast of the underlying tissue in medical segmentation tasks [159] and medical image enhancement [160]. The attention mechanism strengthens the ability of GAN to learn features accurately and improves the accuracy of anomaly location.

4. Applications of GAN-based anomaly detection

Anomaly detection is a crucial problem as well as a basic requirement in machine vision, but detecting unseen anomalies is challenging. Anomaly detection is widely applied in industry, infrastructure, medicine, and other areas, and is playing an increasingly important role in detecting problems. At present, GAN and its adversarial training strategy are used to detect anomalies in these areas, as shown in Fig. 16, and better results have been achieved using these methods than previous methods. GAN skillfully uses the concept of an adversarial game to train discriminators to successfully identify anomalies, but it does not need to know the real distribution of the normal data. These advantages allow GAN to be widely used in detection and recognition tasks.

In this section, the implementation of GAN in these anomaly-detection applications is reviewed. In particular, this section summarizes the key problems that GAN solve in these areas, and expounds the potential research direction of GAN, along with new challenges that need to be solved in the future.

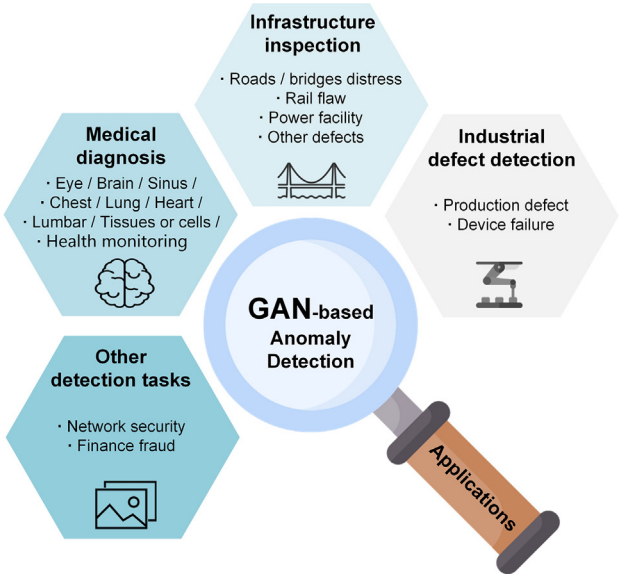


Fig. 16. Main applications of GAN-based anomaly detection.

4.1. Industrial defect detection

In industrial production, machine systems often break down and generate a wide variety of product defects due to their long-time operation and the degradation of components. Defect detection is an essential link to ensure the reliability and stability of systems and control product quality. In recent years, deep learning technologies have been the most successful applications in industrial defect detection. A detector is obtained by training with a large numbers of samples, and the detector can then perform automatic detection tasks such as product defects, abnormal operation of equipment, and fault diagnosis. At present, although DAD has been successfully applied in many industrial detection tasks and can achieve state-of-the-art performance, the task is still challenging because it requires extensive training samples and high sample quality. As identified in the latest industrial detection review [161,162], three key problems in the current DAD tasks can be summarized as follows:

- Cases in which there are few or no real defects, it is difficult for the detector to learn the feature representation accurately, which is very difficult for the data-driven industrial detection task.
- It is difficult to determine the class of defects, and the labeling process of new defects requires the assistance of production experts.
- Industrial production requires real-time and efficient defect detection, which is a routine task that must be solved.

In industrial detection, due to the lack of defects, detection is often not accurate enough to use deep learning to extract the defect features directly from the limited, unbalanced data. GAN can not only generate defect samples to effectively alleviate the problem of data imbalance, but also detect defects by comparing the differences between input and reconstructed samples after the representation learning of normal features. As a result, GANs are being more frequently used in industrial areas. We summarize the application of GAN in industrial defect detection, including industrial product defect detection and device fault diagnosis.

4.1.1. Production defect detection

Due to improper human operation, process design issues, or aging equipment, product defects are inevitable in the process of industrial production. Defects affect the performance of products and result in losses for the companies involved, and therefore quality inspection is an indispensable link in the production process. Reasonable control of the rate of quality products effectively reduces the production cost of companies.

In the visual detection of industrial products, based on the assumption that the latent features of normal samples are different from that of the defects, GAN is used to capture the feature distribution of normal samples in latent space, and the trained model is used to detect defects. Generally, normal samples are only used for training the GAN, and a combination of normal samples and abnormalities are used for testing and verification.

Based on DCGAN [64], Liu et al. [163] introduced a one-class classification network to detect strip steel surface defects and proposed a loss function to improve convergence speed and stability. The one-class classifier can detect defects of different sizes, shapes, and types, and achieve an average detection accuracy of 94% with real-world datasets. Additionally, its structure is similar to that of DCGAN. Lai et al. [164] used Stochastic Gradient Descent (SGD) to optimize the representation learned in latent space and distinguish

the qualified samples from defects based on the reconstructed image. Their method demonstrated that GAN can capture industrial images with arbitrary textures and identify defects effectively.

Liu et al. [165] introduced a simulation method based on GAN to address the problem of limited defect samples in manufacturing. They used the Encoder-Decoder as the simulation network and conducted adversarial training. The conversion of the local defect region is given priority, and the defect-free region is refined by wavelet fusion. To amplify the defect sample, Niu et al. [166] proposed surface defect-generation adversarial network (SDGAN), utilizing GAN to generate defects and detect the commutator cylinder surface defects and showed that the method is robust to uneven and poor lighting conditions. SDGAN includes two generators, trained by normal and abnormal data, respectively. One generator generates abnormal samples from normal samples and the other generates normal data from abnormal data. Additionally, four discriminators are used to distinguish between real and generative samples.

In addition to the defect detection examples discussed above, Di et al. [167] proposed a novel detection method combined Convolutional Autoencoder (CAE) and semi-supervised GAN (SGAN). The CAE acts as an advanced classifier to identify detective regions and feeds into a SoftMax layer to form the discriminator. Skilton et al. [168] proposed a method based on GAN for detecting and labeling defects automatically and achieved the same performance as AnoGAN on benchmark datasets. In the latest research, GAN has also played a key role in welding defect detection [169], solar cell quality inspection [170] and Mura defect detection [171,172].

Table 4 shows the research and detection performance of GAN implementation on defect detection tasks in industrial production. The aforementioned studies demonstrate that GAN contributes to superior performance in image generation, defect detection, and classification.

Part of the reason why GAN plays such a key role in industrial visual detection is that it relies on publicly available industrial defect datasets, as shown in Table 5. Regardless of the kind of industrial product defect detection, it is only necessary to collect enough normal samples for network training and rely on the adaptive improvements of GAN. The advantages of GAN allow it to extract the data feature distribution and generate realistic data. This allows it to effectively expand training datasets and help more industrial applications to solve problems of defect detection.

4.1.2. Device fault detection

Fault detection of industrial devices is another important problem in industrial production. Once the industrial system or equipment fails, it leads to product defects or other accidents. Therefore, to ensure the stable operation of the system, fault detection is crucial. Generally, the fault detection of the system monitors the signals with time series, such as vibration, interference, motions and so on. For the anomaly detection of industrial time series data, the key problems are as follows:

- The fault samples in industrial time series data are infrequent and difficult to collect, suffer from data imbalance problems.
- Because the time series data is easily disturbed in the process of collection, the obtained data often contain noise, which increases the difficulty of fault detection.
- The amount of time series data is large, and the abnormal data have periodic features.

In view of the above key problems, many solutions have been proposed for GAN-based fault detection, and its performance is now better than that of traditional state-of-the-art methods.

Due to the lack of real fault data, Xu et al. [176] proposed a classification model for predicting pipeline leakage and used WGAN to synthesize fault samples to enhance data to solve the problem of data imbalance. The original data was converted into an image for use as input to the GAN, and its validity was judged by observing the quality of the generated images. The classifier based on LSTM learned the time correlation of the data and classified the pipeline state to predict leakage. In addition, Cao et al. [177] introduced a GAN that uses 2D images of the transformed fault diagnosis signal in the time domain, thereby requiring limited data to obtain the fault data and reducing the dependence on labeled data. The results demonstrated that GAN has significant potential for use in few-shot classification of fault diagnosis.

Comparing with the 1D time series data, existing CNN architectures are more suitable for 2D image data. Therefore, the method of converting 1D signals into 2D images has practical applications. For different mechanical system fault detection tasks, the most commonly used datasets are the rolling bearing dataset [185] and sound dataset [186]; however, there are few publicly available industrial time series datasets. The fault-detection tasks based on time series data are more complex and difficult and will require the mixing of multi-sensor monitoring data. In addition, detecting anomalous signals is essential in system management. Zhou et al. [187] presented a radio anomaly detection algorithm based on modified GAN (E-GAN). The existence of anomalies in the spectrogram image, which were converted from radio signals by short-time Fourier transform, could be detected based on the reconstruction error and the discriminator loss. Additionally, Li et al. [188]

Table 4
Research and performance of GAN implementation on production defect detection. (Sv-, Un-, Semi-, W- are the abbreviations of supervised, unsupervised, semi-supervised, and weakly supervised, respectively).

Tasks	Reference	Architecture	Datasets	Performance	Method
Steel surface defect detection	Liu set al. [163]	DCGAN	Images obtained from the Handan Steel Factory	Precision: 94.1% Recall: 93.8% F1-score: 93.9%	Un-
Industrial anomaly detection and one-class classification	Lai et al. [164]	DCGAN	Solar panel dataset Wood texture dataset	Precision: 93.8% Precision: 92.0%	Un-
Automatic surface defect inspection	Liu et al. [165]	WGAN	Button defect dataset Road crack dataset [173] Welds defect dataset [174] Micro surface defect dataset [175]	Precision, Recall, F1-score	Sv-
Defect image generation for improving defect recognition	SDGAN [166]	Two generators and four discriminators	Commutator cylinder surface defect data set without labels	Error rate: 1.8% Relative improvement than CNN-only baseline: 49.4%	Un-
Surface defect classification	CAE-SGAN [167]	GAN, Convolutional AE	Defect image of hot rolled plates Defect image of hot rolled strips Defect image of cold rolled strips	Classification rate: 97.2% Classification rate: 98.2% Classification rate: 96.7%	Semi-
Visual detection of generic defects	Skilton et al. [168]	GAN	MNIST Joint European Torus (JET) dataset	Precision: 83.0% Recall: 89.0% F1-score: 86.0% Accuracy scores: 63.0%	Un-

Table 5
Common industrial datasets used in the reviewed literature.

No	Dataset	References	Image size	Num.	Category	Image	Purpose
1	Wood Defect Database [178]	Silvén et al. (2003)	/	42	Normal, 1000 classes defects and annotations	RGB	Wood defects detection and recognition
2	German DAGM 2007 datasets [179]	Wieler et al. (2007)	512×512	3450	Normal, 10 classes defects and annotations	Gray	Defects detection on various textured backgrounds
3	NEU surface defect database [180]	Song et al. (2013)	200×200	1800	6 classes defects and annotations	RGB	Surface defects classification of the hot-rolled steel strip
4	Micro surface defect database [175]	Song et al. (2013)	640×480	35	2 classes defects	RGB	Surface defect detection of silicon steel strip
5	Oil pollution defect database [181]	Song et al. (2014)	640×480	16	Defects	RGB	
6	Rail surface discrete defects Type-I dataset [182]	Jinrui et al. (2018)	160×1000	67	Defects and annotations	Gray	Rail surface defects Inspection
7	Rail surface discrete defects Type-II dataset [182]		55×1250	128		Gray	
8	MVTec AD Dataset [183]	Jinrui et al. (2019)	700×700 to 1024×1024	5354	Normal, 15 classes defects and pixel-precise annotations	RGB	Industrial texture and defect detection
9	Kolektor Surface-Defect Dataset [184]	Tabernik et al. (2019)	Width: 500 Height: 1240 to 1270	399	Defects and annotations	Gray	Surface defect detection

proposed the Fusing Convolutional Generative Adversarial Encoders (fCGAE) method to create fault-detection models from just the normal signals. Yan et al. [189] selected high-quality synthetic fault data samples with GAN for data augmentation, and Li et al. [190] proposed a semi-supervised FDD approach for the building of the HVAC system based on the modified GAN.

In the development process of industrial internet of things and intelligent manufacturing in the In the development process of industrial Internet of Things and intelligent manufacturing in the future, detection tasks must be closely combined with time series data and image data to achieve high industrial system safety and real-time, on-line anomaly monitoring.

4.2. Infrastructure inspection

Common infrastructure anomalies include road distress, bridge cracks, bridge cables, rail strains, and power transmission failures. The purpose of infrastructure inspections is to keep the infrastructure in question in good condition and prolong its service life, while improving the ability to predict accidents. The automatic inspection of infrastructure is an important research area in the development of intelligent transportation. Compared with traditional laser detection, the development of visual detection methods provides a faster and more feasible solution. In recent years, with the development of artificial intelligence (AI), deep learning has played a key role in object detection, image segmentation, and classification. Many methods have been used to detect faults in infrastructure, but they are also faced with problems such as data imbalances because it is difficult to obtain data.

GAN enables the generator to learn the feature distribution of healthy infrastructure data through adversarial games, so that the generated images are more consistent with the spatial distribution of normal data, and at the same time distinguish the data outside the distribution as anomalies. As a result, GAN provides a viable solution to key challenges in infrastructure inspection tasks.

4.2.1. Distress detection on roads and bridges

As large-scale structural infrastructure, roads, and bridges inevitably experience surface distresses after structural degradation or external damage, they need to be evaluated regularly to ensure traffic safety. Road distresses mainly include cracks, ptholes, and patches, whereas bridge distresses include cracks, spalls, and other damages [191], as well as cable problems. Among them, cracks are the main surface distress in road and bridge maintenance, and are thus key parts of detection tasks.

Cracks are some of the earliest signs of structural degradation. Crack detection is used to evaluate the safety performance of roads and bridges. Traditional image processing techniques rely on the gray difference between the crack and the background to achieve segmentation [192], but it is easily affected by noise and leads to poor generalization ability. Deep learning was first applied to crack detection [193], and it demonstrates a strong generalization ability and robustness in extracting global and sensitive features of crack images. However, the crack detection based on deep learning still faces key problems, as follows:

- The span of roads and bridges is large, and therefore the necessary detection range to cover is wide.
- Cracks include various shapes and types, such as different widths, aspect ratios, and directions [194], and crack features are highly complex.
- Because the crack image is often disturbed by uneven illumination, shadow, motion blur, and the similarity in appearance of crack and road textures, crack images with low contrast can result, making it difficult to segment and extract the crack [195].
- Due to the lack of labeled high-quality crack datasets, there is a problem of data imbalances, which hinders the rapid application of the new technologies.

To address the key problems summarized above, GAN can provide a good strategy for crack detection. GAN can extract feature distributions from a small number of real crack images and generate extensive, realistic crack images to enhance the performance of pixel-level crack detection.

Crack dataset expansion: In general, most of the proposed detection methods have been verified using public crack datasets, but because the environment of image acquisition is relatively simple, it can only achieve detection in simple scenes. Therefore, Mei et al. [196] collected road crack images across a variety of natural scenes and made pixel-level labeled EdmCrack600 datasets. A road-detection solution combined cWGAN and a connected domain graph to propose the ConnCrack algorithm and used DenseNet121 [197] as the feature extractor. This work provided high-quality datasets and application technologies for using in road crack detection. Zhang et al. [198] proposed a self-supervised Cycle-cGAN, which was implemented by training a cycle network to convert the output back to the input. A dual network was constructed with two GANs, one which was trained to transform an image block into a structural block and the other which was

trained to realize the inverse process. The model converts the crack image into a Ground Truth (GT)-like image with a similar structural pattern and is used for crack detection.

Crack feature extraction and segmentation: The CNN is a powerful method for single-pixel semantic segmentation and detection. However, in crack detection, because accurate GT cannot be obtained and the data is unbalanced, the CNN frequently converges to the state wherein all pixels are regarded as background, which is referred to as the "All Black" issue. To solve this problem, Zhang et al. [199] proposed an end-to-end training Crack-Patch-Only (CPO) supervised GAN, which is based on DCGAN and U-Net [200]. It overcomes the "All Black" issue by inputting large crack images into an asymmetric U-shape generator to force the network to generate GT images. Zhai et al. [201] proposed another unsupervised visual detection framework, which uses DCGAN to learn the feature representations of normal surface samples. Based on the sensitivity of the trained discriminator to abnormal regions, the first three convolution layers were used as feature extractors to realize multi-scale feature fusion and segmentation. Their method showed effectiveness and robustness with the Road Crack Database (CrackIT) [202]. Similarly, Gao et al. [203] introduced a road crack segmentation method based on GAN. U-Net and a cross-layer concatenate network were used as generators to effectively avoid the loss of information and the disappearance of gradients, and the segmentation performance was verified on the CFD [173] and AigleRN [204] dataset.

Crack location: To solve the problems of subjective and heavy workloads for manual labeling, and the low contrast between cracks and pavement, Duan et al. [205] proposed an unsupervised learning mapping method. In feature extraction, the CNN connected by eight residual blocks was used as the generator and a 5-layer full convolution network was used as the discriminator, and the cyclic consistency loss was introduced to improve the accuracy of crack location.

The current GAN-based crack-detection methods mainly focus on image data; however, GAN is also used in distress detection using non-image data. For example, Mao et al. [206] innovatively used GAN to detect distress in bridge health monitoring time series data, by converting them into Gramian Angular Field (GAF) images to facilitate network training. In addition, bridge fault detection mostly relies on traditional methods, such as laser scanning and ultrasonic detection, but the research based on deep learning is less common. Based on the data imbalance observed before and after damage had taken place, Lee et al. [207] proposed a data-generation method based on GAN for damage detection. The model can analyze complex data independently, and the generated data can help train damage-detection methods based on deep learning.

Table 6 shows the research and detection performance of GAN implementation for distress-detection tasks pertaining to roads and bridges. The above studies demonstrate that GAN achieves superior performance in image generation, defect segmentation, and location.

In road and bridge distress detection tasks, the proposal of GAN mainly depends on the availability of public crack image datasets. Previous research [196] has identified nine road crack datasets in detail; therefore, more research should pay close attention to the distress data from other domains or complex natural scenes in the future. Additionally, the most effective way to solve the data imbalance problem is to generate more distress images from different scenes and categories based on GAN, and simultaneously accumulate more distress types that the network has not yet encountered. For the problem of background interference in the image, the image background style transfer [209] methods or road damage image synthesis [210] can be used to simulate crack images in various scenes and form an effective dataset, which can effectively improve the detection performance of the current GAN model.

4.2.2. Rail flaw detection

In railway transportation, once the track, catenary, and infrastructure damage and other flaws occur, this can cause serious safety accidents; therefore, it is necessary to regularly check for faults and defects in the railway system. In flaw-detection tasks, the key-detection objects include rail tracks, track fasteners, catenaries, pantographs, bogies, and other key components [211,212]. Due to the danger of railway high voltage circuit, safety issues generally surrounding transportation, and high detection costs, manual detection is not feasible, but a reliable real-time visual monitoring is still required. Therefore, a feasible alternative is to input real-time images captured by video into a trained deep neural network for real-time flaw monitoring. However, this approach still faces several key problems:

- The range of railway requiring maintenance is very large, the labor and time required to maintain that rail is high, and visual detection technologies with high mobility and reliability are necessary.
- The detection task is easily affected by the external environment, as the railway equipment is exposed to the natural environment all year round and the collected images are often polluted by dust and oil, which increase the difficulty of detection.
- The number of railway fault images is limited and difficult to obtain.

Because examples of abnormal data are difficult to obtain because there are few faults, supervised deep learning methods based on labeled training data cannot be used. In contrast, GAN can learn reusable feature representations from extensive unlabeled datasets, which have an advantage in the unsupervised anomaly-detection task of the railway.

In rail fault detection, the catenaries are the key detection component, and bird nests represent a typical anomaly to be detected because these nests can cause catenary tripping, insulator breakdown, and other faults. At present, finding these bird nests mainly relies on manual detection, which is expensive and inefficient. Based on this, Yang et al. [213] proposed a railway catenary anomaly-detection method based on DCGAN to detect bird nests. This network only uses normal catenary images as training data to extract data features from unlabeled data. In addition, GAN is also used to detect anomalies in other railway infrastructure components. Lyu et al. [214] developed an anomaly detection method, referred to as Catenary Supporting Components (CSCs), which combines a CNN and a GAN to judge whether there is a fault and issues an alarm to prevent the occurrence of railway system accidents. This method can correctly judge insulator anomalies and has good fault-detection capabilities.

In addition, the isoelectric line is an important part of the catenary support device of high-speed railways, and it is prone to failure for vehicles that have been driven for a long time. Lyu et al. [215] proposed an isoelectric line fault detection method using GAN, which uses Faster R-CNN to extract features and accurately locate the isoelectric line from an input image, and uses DCGAN to obtain its countermeasure representation of features. Finally, an anomaly classification criterion is used to identify faults in the isoelectric line. Combined with CNN and AnoGAN, Xue et al. [216] developed an unsupervised anomaly detection system to detect turnout current faults in real time.

In addition to image-based rail flaw detection, researchers have also used GAN to detect anomalies in acoustic data. To overcome the noise interference caused by mechanical interactions between

Table 6
Research and performance of GAN implementation on distress detection for roads and bridges.

Tasks	Reference	Architecture	Datasets	Performance	Method
Crack dataset expansion	Mei et al. [196]	cWGAN, DenseNet121 [197]	EdmCrack600 datasets	Precision: 80.9% Recall: 76.6% F1-score: 77.0%	Un-
	Cycle-cGAN [198]	GAN	Four crack data sets	/	Un-
Crack feature extraction and segmentation	Zhang et al. [199]	DCGAN, U-Net	CFD [173]	Precision: 88.0% Recall: 96.1% F1-score: 91.9%	Sv-
			CrackGAN dataset (CGD)	Precision: 87.0% Recall: 96.0% F1-score: 91.3%	
			Dataset [208]	Precision: 86.5% Recall: 94.2% F1-score: 91.3%	
	Zhai et al. [201]	DCGAN	Wood Defect Database [178]	Pixel accuracy: 79.9%	Un-
			Road Crack Database [202]	Pixel accuracy: 58.8%	
	Gao et al. [203]	GAN, U-Net	AigleRN [204]	Precision: 91.5% Recall: 73.4% F1-score: 77.3%	Un-
Crack location	Duan et al. [205]	GAN, U-Net	CFD [173]	Precision: 89.7% Recall: 82.5% F1-score: 85.1%	Un-
	Mao et al. [206]	AE, GAN	Time-series datasets	Detection accuracy: >90.0%	Un-
	Lee et al. [207]	GAN	/	/	Un-

the wheel and the rail, Wang et al. [217] introduced an improved regularized least-square GAN (LSGAN) for acoustic emission signal denoising. The improved LSGAN retains more crack signal details than the traditional denoising method and effectively removes both statistical and mechanical noise.

To address the difficulties of rail flaw detection, one feasible approach is to utilize mature Unmanned Aerial Vehicle (UAV) inspection technology to implement real-time anomaly monitoring along the railway line and continuously accumulate abnormal data. Additionally, the application of UAV can notably reduce workloads and achieve anomaly detection in a wide range of settings.

4.2.3. Power facility detection

To maintain the reliability, availability, and sustainability of power facilities, it is necessary to conduct regular troubleshooting. Manual patrols and helicopter assistance are typically used to conduct necessary maintenance or replacement tasks before any faults have an opportunity to cause major failures, but this solution is inefficient, dangerous, and costly [218]. Power facility detection mainly includes fault detection for power equipment, transmission lines, power towers, and underground pipe corridors, in which the main anomalies include the aging of components and the introduction of foreign bodies.

In recent years, many researchers have implemented automatic visual fault detection of power facilities using UAVs and climbing robots; however, this detection task is faced with high-precision requirements. Based in part on the reviews of visual power-detection methods using deep learning that was conducted by Jenssen et al. [219] and Liu et al. [220], the key problems of deep learning in power facility DAD are summarized as follows:

- Power facility detection needs long-term maintenance.
- The class of faults are unbalanced, wherein some anomalies such as broken insulators and cracked poles are relatively rare.
- Most of the captured images contain complex natural scenes wherein the object is difficult to separate from the background, and most of the detected objects are small and low in contrast, which makes it difficult to identify anomalies and causes poor results.
- Lack of public benchmark datasets.

In view of the above key problems, such as data insufficiency and poor data quality, the current GAN methods have demonstrated the advantages of advances in image generation and detection and provide a feasible approach for the future.

Research based on GAN alleviates the data problem associated with power facility detection task at different levels. Among them, the conventional detection task is for the detection of power transmission line components, including image resolution enhancement, object detection from complex backgrounds, and image segmentation and generation.

Image resolution enhancement: Electrical insulators are widely used in power facilities and are easily damaged, implying they need to be maintained frequently. Traditional deep learning provides a safe and fast method for the detection of this damage, but requires a large amount of training data. To effectively generate high-quality images, Luo et al. [221] designed a new model - Balanced and Progressive GAN (BPGAN) to embed class information in training to generate high-resolution images, which improved the performance when distinguishing between normal and abnormal samples. The results demonstrate that this is an effective method to learn data representations in a stepwise manner from low to high resolution.

Object detection in complex background: To realize the detection of electric wire in cluttered environments, Chang et al. [222] proposed a real-time recognition and segmentation method for overhead ground wire based on cGAN, in which the generator used a skip-connected end-to-end CNN and the discriminator used a multi-level CNN. During adversarial training, the learning of the generator combined the multi-scale feature learning for straight line and stripe features and the discriminator learned the loss function to sharpen the edges of the saliency maps by inputting the wire image obtained by aerial photography into cGAN and outputting a labeled image that contained only the wire.

Improving the detection accuracy from pixel-level segmentation: Because of the significant difference in color and shape between insulators and the effects of a chaotic background, the segmentation of insulators is still very difficult. To address this challenge, Chang et al. [223] applied cGAN to detect insulators of different sizes, materials, and shapes and achieved pixel-level segmentation. They used a training strategy for the network wherein it was trained in stages, first training the coarse position in the labeled samples for object location, and then inputting the finely

labeled segmented samples into the same model for training. However, insulator images that included Gaussian noise and different transparency were also used for model training to enhance detection performance. Among them, the location of the labeled samples can help the network converge in the direction of fine segmentation and improve the performance of detection and segmentation.

Improving the detection accuracy from generating multi-scale and multi-background data: It is very difficult to collect a wide variety of samples for insulator detection; therefore, Chang et al. [224] proposed an insulator image-generation method based on cGAN, which considered the expansion of the dataset with respect to the diversity of image backgrounds and insulators. Using an adversarial training framework, three end-to-end segmented networks were used as generators for performance comparison. Additionally, the training strategy of modifying super-parameters in stages was adopted to accelerate the convergence of the model.

GAN has also been used for intrusion detection in the next-generation smart electrical grid. Siniosoglou et al. [225] proposed a novel Autoencoder-Generative Adversarial Network (GAN) architecture for detecting operational anomalies and classifying Modbus/Transmission Control Protocol (TCP) and Distributed Network Protocol 3 (DNP3) cyberattacks and is validated in four real electrical grids environments. Additionally, improved GAN [226] solves the problems of insufficient faulty samples for faulty line detection.

Table 7 shows the research and detection performance of GAN implementation on anomaly-detection tasks for power facilities. The aforementioned studies demonstrate that GAN is capable of superior performance in image generation and anomaly detection.

Because there is no open dataset of power line facility detection, it is difficult to evaluate the performance of DAD systems. In the current research on fault detection of power facilities using GAN, GAN is used to generate higher-resolution images, but also to seek out methods to obtain multi-scale and multi-scene object images to improve the generalization ability of the model, such as in image style transfer [209]to improve the detection accuracy. These two methods focus on solving the problems associated with small amounts of data and class imbalances through data enhancement and expansion.

4.2.4. Other infrastructure inspection

In addition to large-scale infrastructure, such as roads, bridges, railways, power facilities, anomaly detection is also successful in the areas of security and transportation. Therefore, GAN can also be applied to these detection tasks.

Security defense: It is a dull yet important task to detect abnormal events in the field of security, such as in the context of indoor security patrol and outdoor behavior abnormal monitoring. There are extensive video data that at first seem unimportant in these applications, and therefore real-time anomaly detection based on data stream is the focus of research in this domain. In security monitoring, Lawson et al. [227] worked with GAN and proposed an anomaly-detection system based on automatic robots to conduct patrol tasks, which was able to recognize anomalies by comparing the current capture view with the normal environment. In addition, abnormal crowd behavior detection is becoming more important in video surveillance scenes. Ravanbakhsh et al. [76] proposed a crowd anomaly-detection method based on GAN, and were the first to apply adversarial training strategy to discriminative tasks. Other research works [228–230,75] have also investigated detecting abnormal targets, behavior, or events in crowded scenes and demonstrated that they can effectively detect abnormal events in surveillance videos.

Transportation anomaly detection: Transportation safety is another research topic in the domain of anomaly detection. GANomaly [80] used cGAN to jointly learn the generation of high-dimensional image space and latent space for anomaly detection. They applied the Encoder-Decoder-Encoder model to learn the data distribution of normal samples, wherein the larger distance between the generated image and the latent vector was used to measure the anomalies of the new data distribution. To solve the problem of limited samples, Akçay et al. [148] proposed an unsupervised anomaly-detection model, which innovatively employed an Encoder-Decoder CNN with skip connections to thoroughly capture the multiscale distribution of the normal data in the image space.

Traffic anomaly detection: In the event of mechanical failure or external interference during vehicle driving, this can cause serious accidents, so abnormal early warnings and emergencies are necessary. Sun et al. [231] proposed a fault-prediction method using GAN to simulate the anomaly-detection process conducted by scene experts, which was verified using Isuzu vehicle data. In addition, to detect vehicle deviations from the normal driving track more accurately, Qiu et al. [232] developed a multi-pattern recognition system to obtain vehicle information. The system relies on GAN to measure the score difference between the predicted and actual signals to accomplish unsupervised driving anomaly detection.

Environmental monitoring: Automated sewer defect detection has become an important focus for improved management and maintenance of urban sewer systems. Situ et al. [233] adopted styleGANs to efficiently produce high-quality images with various styles and high-level details pertaining to multiple different types of sewer defects to solve the issue of insufficient data and unbalanced samples. Additionally, anomaly detection using hyperspectral images (HSIs) has received significant attention in environmental monitoring [234]. Jiang et al. [105] proposed a weakly supervised spectral constrained GAN for hyperspectral anomaly detection (HAD). A discriminative end-to-end reconstruction, with the background homogenized and anomalies made salient, can be obtained, although the anomalous samples are limited

Table 7
Research and performance of GAN implementation on power facility detection.

Tasks	Reference	Architecture	Datasets	Performance	Method
Image resolution enhancement	BPGAN [221]	BGAN	Raw insulator images	Recognition Rate: 79.5%	Un-
Object detection under complex background	Chang et al. [222]	cGAN	Datasets contain wire with clear strip texture (Set1 and Set2)	Detection accuracy: 96.6% and 93.0%	Un-
Improving the detection accuracy from pixel-level segmentation	Chang et al. [223]	cGAN	Images contained insulators of different scales, materials, and shapes	The intersection over union (IOU): 82.2%	Un-
Improving the detection accuracy from generating multi-scale and multi-background data	Chang et al. [224]	cGAN	Real-world insulators images, Massive aerial images from Unmanned Aerial Vehicle (UAV)	/	Un-

and can appear similar to the background. Lastly, anomaly detection in astronomical surveys [108] has recently been introduced and seems an interesting area for future research.

4.3. Medical diagnosis

Medical diagnosis is one of the most popular applications of DAD. Anomaly recognition for medical images can help distinguish the grade of diseases presented in the images and allow medical personnel to implement timely medical interventions. For example, identifying lesions from medical images is necessary for diagnosis, treatment, and prognosis. The most common medical data are images, such as X-ray, Magnetic Resonance (MR), and Optical Coherence Tomography (OCT). At present, medical research mainly focuses on medical image analysis and diagnosis.

Medical image analysis used traditional image processing algorithms in its early stages, then advanced to combining the feature engineering of supervised learning, and then to feature learning with CNNs, which is widely used at present. Nowadays, with its strong feature learning ability, deep learning has gradually become a key technical method in medical image analysis, especially for anomaly-detection tasks. In conjunction with several recent reviews [235–237], we have concluded that there are still several key problems in medical diagnosis:

- There is a scarcity of available labeled images, and the task of annotating lesion areas is arduous. The scarcity of images can be attributed to the time-consuming nature of medical image acquisition and labeling. Generally, labeling experts must have professional medical or related knowledge. Additionally, the diseases being investigated are usually multi-stage, further complicating.
- Due to the rarity of some diseases, medical data often have the problem of class imbalances. Models trained by the dataset corresponding to one or more diseases cannot detect other undiscovered diseases, which limits the application of these technologies in disease screening.
- Many popular deep learning algorithms are not specially developed to solve the problem of interpretability in the medical diagnosis. Therefore, it is necessary to introduce clinical medical experts to evaluate the interpretation of different diseases, such as human-in-the-loop learning, which can be used to design interpretable clinical diagnosis networks to imitate the decision-making process of medical experts and avoid misdiagnosis.

Although labeled data may be scarce in the medical field, unlabeled medical data are easily available. In this context, GAN has initiated a new unsupervised anomaly-detection paradigm, one that does not need any prior pathological data. The basic method is to model the distribution of the healthy image through the generator and reconstruct the generated image which is most similar to the training image. The discriminator identifies lesions by learning the probability distribution of the training images that describe the healthy pathology. Any images that do not belong to this distribution are regarded as abnormal, that is, abnormalities are detected from the differences between the input and reconstructed images.

4.3.1. Disease diagnosis and health surveillance

With the ability to synthesize realistic images from latent distributions following real data distribution, GAN has been widely used in medical diagnosis [235], among which the key applications are disease prediction and lesion region detection. Fig. 17 shows examples of the successful applications of GAN in medical diagnosis. The following subsection introduces the research progress of GAN with respect to time and specific parts of the anatomy.

Eye: The prospective anomaly-detection method, AnoGAN has demonstrated that GAN can effectively detect abnormalities in retinal OCT images. The main idea is to learn, in an unsupervised manner, the manifolds of normal anatomical changes based on DCGAN and well-designed loss functions, and rely on anomaly scores to identify anomalies. Since this original implementation, this work has inspired much research in medical diagnosis. For example, by combining WGAN and AE, f-AnoGAN has been proposed to identify disease images quickly using unsupervised learning. It establishes a model for generating health training data and detects anomalies based on anomaly scores.

Brain: Due to the differences in size, shape, contrast, and location of brain injuries, the automatic detection of brain injuries is an important and challenging clinical diagnosis task [244]. Since the success of AnoGAN, researchers [245,246] have introduced the idea of adversarial games into the anomaly detection of brain MR images.

Chen et al. [247] used an adversarial AE to understand the data distribution of healthy brain MR images and highlighted lesions by calculating the residuals of lesions and reconstructed images. ANT-GAN [239] used two generators and two discriminators to train the model, forcing the model to generate normal samples when abnormal samples were input, and abnormal samples when normal samples are input. The lesion could then be annotated through the forced conversion between abnormal and normal samples. SegAN [238] was introduced for anomaly segmentation of brain images. It used a full-CNN as a segmentor to generate a label graph and an adversarial critic network. The multi-scale L1 loss function was proposed to force the critic and the segmentor to learn to capture the global and local features of the pixel-to-pixel long-distance and short-distance spatial relationships. Rezaei et al. [248] proposed an adversarial training framework to segment brain tumors. This work used patient-wise full-CNN as a segmentor to generate label maps, while the discriminator could distinguish the results from the ground truth and segmentor.

In addition, based on a large number of healthy images, Han et al. [249] designed a two-step method of brain MRI slice reconstruction based on GAN to detect different stages of Alzheimer's disease. It reconstructs the brain image to detect outliers in the learning feature space or from the highly reconstructed loss. Similarly, VA-GAN [250] has been proposed to detect Alzheimer's disease.

Sinus: In minimally invasive surgery, such as endoscopic surgery, a limited visual field and low depth perception limits the surgeon's ability to employ eye-hand coordination and provides limited site information. This can lead to accidental damage to important structures and sub-optimal surgical results. Therefore, Lin et al. [251] used LC-GAN (based on CycleGAN) to segment surgical instruments from images provided during paranasal sinus surgery. This technique provides guaranteed automatic segmentation and tracking of surgical instruments and improves performance for minimally invasive surgeries.

Chest: To detect and locate chest abnormalities, Wolleb et al. [240] proposed DeScarGAN to detect chest pathological changes, such as pleural changes caused by brain atrophy or pleural effusion. They also introduced a novel disease-specific architecture with skip connections, a splitting of the networks into weight-sharing subnetworks, and an identity loss to function as an identity-preserving mechanism. Swiecicki et al. [252] used only images that did not contain cancer to train an inpainting GAN for cancer detection, and the network completed the removed part during inference. A significant error during the completion of an image part was considered an indication that such a location is unexpected and thus abnormal.

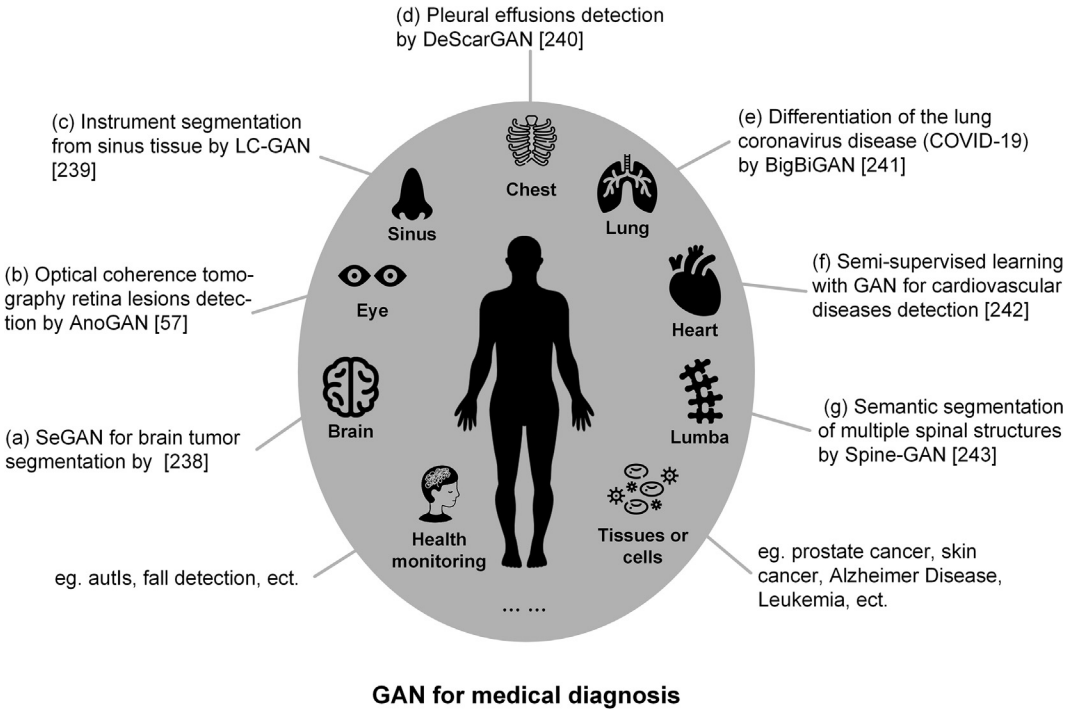


Fig. 17. Applications of GAN-based for medical diagnosis. (a) Medical image segmentation (Xue et al. 2018) [238], (b) Abnormality detection of optical coherence tomography images of the retina (Schlegl et al. 2017) [56], (c) Image translation based on GAN for endoscopic images (Lin et al.2020) [239], (d) Pleural effusions detection (Wolleb et al.2020) [240], (e) Differentiation of the coronavirus disease (Song et al.2020) [241], (f) Cardiovascular diseases detection (Madani et al.2020) [242], (g) Semantic segmentation of multiple spinal structures (Han et al.2018) [243].

Lung: The outbreak of novel coronavirus (COVID-19) in 2019 was significant, and it is a virus that can cause viral pneumonia. One of the effective ways to detect such lung diseases is to use X-rays to scan the lungs and identify the lesions. However, COVID-19 detection tasks based on deep learning face a challenge - the lack of chest X-ray image datasets for COVID-19 [107]. Therefore, researchers have conducted research on data enhancement. For example, Loey et al. [253] used traditional GAN to detect COVID-19 in chest X-ray images. The training data included COVID cases, normal images, pneumonia bacterial images, and pneumonia virus images, and it relied on multiple deep transfer learning models to detect the virus from X-ray images. Additionally, Loey et al. [254] used the data enhancement afforded by affine image transformations, combined with cGAN, to help detect COVID-19. Similarly, Khalifa et al. [255] implemented deep transfer learning and image enhancement based on limited datasets, which solved the problem of data over-fitting. In addition, Song et al. [241] designed an end-to-end representation learning method based on BigBiGAN that trained a linear classifier to detect COVID-19 from suspected patients in real time. One particularly creative research endeavor [256] proposed a Federated Differentially Private Generative Adversarial Network (FedDPGAN) to detect COVID-19 pneumonia for sustainable smart cities.

Heart: The lack of significant labeled data makes it difficult for trained classifiers to maintain high discrimination performance on new or unseen cardiology image datasets. Madani et al. [242] used a semi-supervised learning GAN to solve the problem of lack of labeled data and data domain over-fitting, and demonstrated that the semi-supervised learning GAN required less data than a traditional supervised CNN. In addition, Zhang et al. [257] proposed Semi-Coupled-GAN (SCGAN) to detect left ventricular coverage in cardiac MR images. It used two generators to learn normal and abnormal features and generated corresponding antagonistic samples, but used the same discriminator. Through the identification of missing basal and apical slices in a cardiac MR volume, the model was able to accurately evaluate cardiac volume and ultimately assist in cardiovascular disease diagnosis.

Lumbar: Finding abnormalities and diagnosing diseases from spinal MRIs is an effective clinicopathological diagnosis method at present, but there are some challenges in this work, such as the simultaneous segmentation of multiple disease structures, structural diversity, and weak differences. Han et al. [243] proposed Spine-GAN to automatically segment and classify the intervertebral disc, vertebral, and nerve foramen at the same time. The segmentation network used an atrous convolution autoencoder module for spinal image representation and pixel-level classification, and creatively embedded a LSTM module between the encoder and the decoder, which allowed it to dynamically model the spatial pathological correlation between normal and abnormal spinal structures.

Tissues or cells: In the detection of abnormal cells or tissues, many methods have been developed to analyze and detect different diseases. Sparse-GAN [91] has been successfully used for disease screening of Diabetic Macular Edema (DME) and Choroidal Neovascularization (CNV) datasets. It trains using only the healthy patient data and uses the anomaly activation map of the lesions to explain the detection results and achieves abnormal disease detection during the latent period. Aida et al. [258] used cGAN to segment cancer stem cells in phase contrast imaging. In addition, for the detection of acute lymphoblastic leukemia, Tuba et al. [259] first introduced the idea of Generative Adversarial Optimization (GAO), the goal of which is to divide lymphocytes into normal cells or blasts. Kohl et al. [260] utilized an adversarial network that discriminates between expert and generated annotations to train fully convolutional networks for semantic segmentation, and the network was able to learn from the MRI images of patients to segment invasive prostate cancer. For the detection of skin lesions, Udrea et al. [261] developed an algorithm for the diagnosis of skin cancer

based on cGAN. In the aforementioned studies of abnormality detection in cells or tissues using GAN, the goal is to find potential diseases from the obtained images and to use them for clinical detection.

Health monitoring: Identifying abnormalities in human behavior is an efficient way to monitor health and detect diseases early. For example, child vocalizations can be used as an objective sign to identify developmental disorders, such as autism. Deng et al. [262] first classified the speech of children affected by developmental disorders through the feature representation learning of GAN. In addition, Nho et al. [106] proposed an automatic fall-detection method that was applicable for a wearable device, which incorporated GAN-based anomaly detection that was partially surrounded with User Initial information features. The technology was designed to promptly alert caregivers when a fall was detected to reduce the likelihood of injuries for elderly people. These studies demonstrate that the detection of behavioral abnormalities will also be an essential item for medical diagnosis in the future.

4.3.2. Implementation performance

The purpose of this chapter is to help researchers better understand the applications and innovations of GAN in medical diagnosis. Table 8 shows the research and detection performance of GAN implementations on different anatomical disease diagnosis tasks. The aforementioned studies demonstrate that GAN shows superior performance in image generation and segmentation, which makes it a good artificial assistant technology for medical diagnosis with high accuracy. However, at present, GAN is difficult to use as a reliable technology in medical image analysis. This is because there are unexplainable problems in the reliability of the generated data and detected results in clinical application [235], that is, the reliability of generated and detected results has not been accurately evaluated. This is the main obstacle to the practical application of GANs and other deep learning methods in medical environments.

A previous medical imaging review [237] based on GAN has identified 54 publicly available medical MR image datasets, including more than a dozen anatomical datasets (listed above), which provide convenient datasets for medical diagnosis research. In addition, the University of Freiburg also provides segmentation datasets for transmission electron microscopic cell recordings datasets [266], which contain hundreds of images of human and mouse cells. Furthermore, Hernandez-Matas et al. [267] collated and shared 18 publicly available retinal image datasets that can be used in the diagnosis of different retinal diseases. The more medical image datasets are made public, the easier it is to verify GAN-based anomaly-detection methods.

4.4. Other detection tasks

In addition to the three main applications mentioned above, GAN is widely used for anomaly detection in network security and financial transactions.

4.4.1. Network security detection

Network systems around the world are attacked and controlled by malicious activities every day, which can lead to serious problems, such as the system being disabled or even completely collapsed. Therefore, anomaly detection has generated an extensive research in network security. Among them, network intrusion is the most common. To detect attacks on the Internet of Things (IoT), Li et al. [268] constructed a GAN with LSTM to detect the abnormal behavior of cyber-physical systems based on the time series. Because most of the time series data are regular, this systems uses an RNN to construct the GAN network. Lin et al. [269] proposed a novel GAN framework for adversarial attack generation against the intrusion-detection system and enhanced the robustness of the system to hostile attacks. Additionally, GAN-based Intrusion Detection System (GIDS) [270] have been proposed to detect known attack signatures.

It is challenging to identify OOD samples in task-oriented dialogue systems. Therefore, OodGAN [271] has been proposed for OOD data generation, and Zeng et al. [272] proposed an efficient adversarial attack mechanism to detect OOD samples. The system log can be used to track the state of the system and record valuable information, so LogGAN [273,274] has been proposed to detect anomalies in those variables. Similarly, Rigaki et al. [275] and Amin et al. [276] also used GAN to improve the defense against malware attacks in mobile networks.

4.4.2. Financial fraud detection

Fraudulent activities in financial transactions include telecommunications frauds, abnormal stock price manipulations, and counterfeit banknotes. In general, the data that are used to build an anomaly-detection model usually include several contents, such as the user ID, the amount spent, the transaction time, and the geographic location. Distinguishing anomalies from these specific time series data is a common fraud-detection method. For telecom frauds, Zheng et al. [277] used a GAN to calculate the fraud probability of each large transfer to identify frauds, and Sethia et al. [278] designed multiple adversarial networks to generate pseudo credit card data to improve the efficiency of their network, which solved the problem of high imbalances of credit card transaction data. Herr et al. [279] introduced variational quantum–classical Wasserstein GANs (WGANs) to address training instabilities and embedded the model in a classical machine learning framework for credit card fraud detection. Chen et al. [280] combined sparse AE to learn the representation of fraud-free data and train a one-class GAN to detect frauds in trading markets.

In addition, GAN is effective, even when only using normal data to train the network, to learn normal stock market behavior and detect abnormal trading behavior caused by stock price manipulation [281]. For the task of identifying the authenticity of banknotes [282], a supervised GAN has also been used to predict counterfeit and genuine banknotes.

5. Discussion

GAN was originally proposed to produce plausible synthetic images and has achieved exciting performance in the computer vision. Although GAN is widely used in anomaly detection, there are many challenges, and it still has broad prospects for future applications. We outline open research issues and challenges faced while adopting GAN-based anomaly-detection techniques for real-world problems.

5.1. Outstanding issues

As a generation model, compared with other models, such as Boltzmann machines [283] and Generative Stochastic Networks (GSNs) [284], GAN only uses back propagation and does not need a complex Markov chain, so it can be widely used in unsupervised and semi-supervised learning. However, GAN-based anomaly detection still faces three major internal and external challenges.

5.1.1. Internal: Sensitivity and specificity of representation

For anomaly detection tasks, the goal is that the representations learned by GAN should be compact to ensure it is difficult to represent abnormal samples while accurately representing normal samples. However, too strict constraints on the latent space will

Table 8
Research and performance of GAN implementation on different anatomy disease diagnosis. (SD-OCT: Spectral-domain optical coherence tomography, BraTS: Brain Tumor Image Segmentation, LiTS: Liver Tumor Segmentation, NCC: Normalised Cross Correlation, mDSC: mean Dice Similarity Coefficient, mIoU: Intersection over Union, MSE: Mean Square Error, OCT: Optical coherence tomography, ALL: Acute Lymphoblastic Leukemia, FCN: Fully Convolutional Networks, CPESD: Child Pathological and Emotional Speech Database)

Anatomy	Reference	Architecture	Datasets	Performance	Method	Code
Eye	AnoGAN [56]	DCGAN	SD-OCT	AUC = 0.890	Un-	√
	f-AnoGAN [86]	WGAN		AUC = 0.930	Un-	√
Brain	Chen et al. [247]	WGAN-GP, AAE	BraTS 2015	AUC = 0.923	Un-	/
	ANT-GAN [239]	GAN	BraTS2018/ LiTS dataset	Precision: 0.924/ 0.772, Recall: 0.910/ 0.779, F1-Measure: 0.917/ 0.776	Un-	/
	SegAN [238]	GAN, U-net, skip connections	BraTS 2013 Leaderboard/ BraTS 2015	Dice score: 0.840/ 0.850, Precision: 0.700/ 0.700, Sensitivity: 0.650/ 0.660	Sv-	/
	Rezaei et al. [248]	cGAN, CNN	BraTS 2017	Dice score: 0.680, Sensitivity: 0.990, Specificity: 0.980	Semi-	/
	Han et al. [249]	GAN, U-Net	OASIS-3 Dataset	AUC: 0:917	Un-	/
	VA-GAN [250]	WGAN	Unknown synthetic dataset	Mean NCC scores: 0.940	W-	√
Sinus	LC-GAN [251]	CycleGAN	Sinus surgery dataset	15.1%∼19.4% better mDSC and 17.9%∼ 22.9% better mIoU than CycleGAN.	Sv-	/
Chest	DeScarGAN [240]	GAN	Unknown synthetic dataset	Dice score: 0.853, AUC: 0.988	W-	√
	Swiecicki et al. [252]	WGAN	Digital breast tomosynthesis dataset	/	Un-	/
Lung	Loey et al. [253]	GAN, Alexnet, Googlenet, Resnet18	COVID-19 dataset	Precision, Recall, F1 Score, Testing Accuracy	Sv-	/
	Loey et al. [254]	cGAN, AlexNet, VGGNet16/19, GoogleNet, ResNet50	COVID-CT-Dataset	Sensitivity, Specificity, Precision, Accuracy, F1 Score Balanced accuracy	Sv-	/
	Khalifa et al. [255]	GAN, AlexNet, SqueezeNet, GoogleNet, ResNet18	Pneumonia data set	Precision, Recall, F1 Score, Total Accuracy	Sv-	/
	Song et al. [241]	BigBiGAN	Patient CT images	AUC: 0.850, Sensitivity: 80.0%, Specificity: 75.0%	Sv-	√
	FedDPGAN [256]	GAN, Resnet	Covid-chestxray-dataset	Accuracy: >91.9%	Un-	/
Heart	Madani et al. [242]	GAN	NIH PLCO dataset/ NIH Chest X-ray	Accuracy: 80.0%/ 93.7%	Semi-	/
	SCGAN [257]	GAN	UK Biobank dataset: Missing Apical Slices (MAS)/ Missing Basal Slices (MBS)	Accuracy: 92.5%±0.5%/ 89.3%±0.4%, Precision: 87.6%±0.4%/ 89.1%±0.3%, Rcall rate: 90.5%±0.5% / 91.7%±0.4%	Semi-	/
Lumbar	Spine-GAN [243]	GAN, Convolution Autoencoder, LSTM	Neural foramen/ intervertebral discs/ lumbar vertebrae	Pixel accuracy: 0.962±0.003, Dice coefficient: 0.871±0.004 Specificity: 0.891±0.017, Sensitivity: 0.860±0.025	Sv-	/
Tissues or cells	Sparse-GAN [91]	GAN, LSTM	OCT images	AUC: 0.925, Accuracy: 0.841, Sensitivity: 0.951	Un-	/
	Aida et al. [258]	cGAN, CNN	Special cell images	Better segmented using CNN with an adversarial loss model than the CNN-only model	Sv-	/
	Tuba et al. [259]	GAN	ALL-IDB2 dataset	Accuracy: 93.8%, Specificity: 91.2%, Sensitivity: 96.2%	Sv-	/
	Kohl et al. [260]	GAN, U-Net, FCN	MRI dataset	Dice coefficient: 0.410±0.280, Specificity: 0.550±0.360, Sensitivity: 0.980±0.140	Sv-	/
	Udrea et al. [261]	cGAN, U-Net	Clinical images of skin lesions database	With Rotation/ epoch 200: 91.4%	Un-	/
Health monitoring	Deng et al. [262]	GAN, SVM	CPESD database	Unweighted average recall: 44.1% and 46.9%	Un-	/
	Nho et al. [106]	Skip-GAN [148]	HIFD [263]	Accuracy: 96.9%, Sensitivity: 98.2%, Specificity: 95.9%	Un-	/
			MobiFall [264]	Accuracy: 99.9%, Sensitivity: 99.8%, Specificity: 100.0%		
			MobiAct [265]	Accuracy: 99.5%, Sensitivity: 99.4%, Specificity: 99.4%		

lead to over-fitting, which affects the discriminating ability of the model for novel and adversarial samples.

In general, the training of GAN requires ensuring that a limited number of representations can generate as many samples as possible, while avoiding mode collapse. However, for anomaly detection, we expect the diversity of generation to be limited to the range of normal samples, which looks like another form of mode collapse. Therefore, proper representation learning is critical. In the absence of sufficient abnormal samples, the decision boundary is often difficult to determine. In this case, the selection of representation considerably affects the shape of the normal and abnormal distributions, thus affecting the sensitivity and specificity of

the anomaly-detection model. Therefore, how to choose an appropriate representation learning method to achieve a balance between sensitivity and specificity is an important challenge faced by GAN-based anomaly-detection methods.

5.1.2. External-I: Generative models vs discriminative models

Compared with generative models, discriminative models have the following three advantages:

- Discriminative models do not need to pay attention to pixel-level details, and therefore they can learn abstract semantic representations better.
- Discriminative models can be compatible with more abundant self-supervised learning pretext tasks, whereas most discriminative/contrasting pretext tasks cannot be directly applied to the training of generative models.
- Most discriminative models can normalize the anomaly scores into probability scores through post-processing, which have better interpretability for the degree of anomaly.

Therefore, many current generative models lag behind the discriminative models in performance, as shown in Table 9 and Table 10. However, the generative models still have the following advantages:

- The attention to pixel details makes it easier to locate anomalies using generative models.
- The generative models can map samples to the semantic distribution in the latent space, which has better interpretability for the types and meanings of anomaly.
- There is no need for complex post-processing to generate the abnormal scores for most generative models, which affords speed advantages.

In conclusion, generative models and discriminative models each have their own advantages. How to develop the GAN-based anomaly-detection methods, and how to learn from the advantages of the discriminative models to overcome problems of weak performance, are important challenges for future researchers.

Table 9
Image-level anomaly detection AUC ranking for Cifar-10 dataset. (Dis-Discriminative, Gen-Generative)

Methods	Mean AUC	Models		Pre-trained
		Dis	Gen	
MSC [285]	0.975	√		√
PANDA [286]	0.962	√		√
CSI [104]	0.943	√		×
DUIAD [287]	0.926	√		×
DN2 [288]	0.925	√		√
DisAug CLR [289]	0.925	√		×
IGD [290]	0.913		√	×
AD in IP [291]	0.906		√	×
USSL [100]	0.901	√		×
SSD [292]	0.900	√		×
UTAD [293]	0.884		√	×
GOAD [125]	0.882	√		×
GEOM [35]	0.860	√		×
RotNet [118]	0.833	√		×
ESAD [294]	0.788		√	×
CAVGA [94]	0.737		√	×
Puzzle-AE [130]	0.725		√	×
JRLFM [127]	0.707		√	×
CutPaste [62]	0.694	√		×
OLED [295]	0.668		√	×
OCGAN [92]	0.657		√	×

5.1.3. External-II: Pre-trained models vs non-pre-trained models

Because there are often few, or even no, abnormal samples, efficient representation learning has always been a difficult task in anomaly detection. In the absence of sufficient training samples, most of the representations learned by semi-supervised, weakly supervised, self-supervised, and unsupervised learning algorithms struggle to match the richness and effectiveness of the representations learned by supervised learning algorithms. Therefore, many scholars attempt to bypass the problems of representation learning and directly use pre-trained models, such as VGG [310], ResNet [311], and EfficientNet [312], to obtain the representation of samples. Pre-trained models based on large data sets such as ImageNet [313] have proven success in transfer learning tasks. Particularly in the field of anomaly detection, they have the following advantages:

- The compatibility of the network architecture of the pre-trained model ensures that it can be applied to various anomaly-detection tasks instantly.
- The pre-trained model ensures that the extracted representation embedding is predictable, separable, can be clustered, and is reasonably interpretable, thus reducing the difficulty of anomaly detection and localization.
- The pre-trained model significantly reduces the complexity and training time of the learning algorithm, reduces the cost of model training, and reduces the difficulty of deployment.

Because the embedding of representation can be easily extracted from pre-trained models, researchers can construct the feature pyramid structure to compare the similarity, to better detect and locate anomalies [299,309,62,302,303]. Pre-trained models have demonstrated their superiority on multiple data sets. Taking the MVTec AD and Cifar-10 datasets as examples, as shown in Tables 9–11, pre-trained models account for the majority of the leaderboards and significantly outperform GAN-based models.

Pre-trained models are equivalent to using extra data sets for learning, which is undoubtedly an unfair advantage over non-pre-trained models. However, the convenience and high efficiency of the pre-trained model are worth popularizing in the field of anomaly detection. Researchers should try to innovate and consider using the pre-trained model to improve the performance of anomaly detection based on GAN.

5.2. Future directions

Researchers generally believe that designing methods to make machines learn in an unsupervised manner directly from unprocessed and unlabeled data is a key problem to be solved at present. Deeper theoretical research is worth studying, regardless of which visual inspection area GAN is applied to in the future, and thus research should focus on explainable learning and data-driven sustainability to meet the higher requirements of DAD tasks.

5.2.1. Zero-shot learning

Zero-Shot Learning (ZSL) [314] refers to using data from the training set to train the model, so that the model can classify objects in the test set, wherein the test set contains categories that were not present in the training set. The ZSL method faces two types of test sets: the first, in which all the test sets are new categories; the second, in which the test sets contain both new categories and categories that were in the training set. The second type is more difficult than the first because the learned models tend to conservatively ascribe new categories to existing ones. Anomaly detection, especially the one-class classification anomaly

Table 10
Image-level anomaly detection AUC ranking for MVTec AD dataset. (Dis-Discriminative, Gen-Generative)

Methods	Mean AUC	Models		Pre-trained
		Dis	Gen	
PatchCore [296]	0.991	√		√
CS-Flow [63]	0.987		√	√
CFLOW-AD [297]	0.982		√	√
DRAEM [298]	0.980	√		√
PaDiM [299]	0.979	√		√
NSA [300]	0.972		√	×
CutPaste(finetune) [62]	0.971	√		√
AnoSeg [132]	0.960		√	×
InTra [301]	0.959		√	×
Gaussian-AD [302]	0.958	√		√
STPM [303]	0.955	√		√
CutPaste [62]	0.952	√		×
DifferNet [304]	0.949	√		√
DFR [156]	0.938	√		√
IGD [290]	0.934	√	√	×
Patch SVDD [305]	0.921	√		×
RIAD [306]	0.917		√	×
FCDD(unsupervised) [38]	0.880		√	×
MOCCA [307]	0.875	√		×
MSC [285]	0.872	√		√
SPADE [308]	0.855	√		√
ARnet [126]	0.839		√	×
CAVGA [94]	0.780		√	×
Puzzle-AE [130]	0.776		√	×
GANomaly [80]	0.762		√	×
LSA [58]	0.730		√	×
GEOM [35]	0.672	√		×
AnoGAN [56]	0.550		√	×

detection task, resembles this challenging situation, which can be considered as a special case of ZSL.

Introducing ZSL into the field of anomaly detection has several benefits:

- The establishment of the hierarchical representation structure of ZSL method helps eliminate biases caused by data imbalances.
- The semantic organization and identification of the ZSL method is helpful to distinguish between different types of anomalies and facilitate downstream tasks.
- Suitable ZSL methods can make the transfer of learned knowledge between different anomaly-detection tasks more convenient.

In previous studies, anomaly-detection techniques were developed independently of ZSL. Currently, only a few ZSL techniques have been introduced into the field of anomaly detection. For example, there have been investigations into fault diagnosis based on ZSL [315] and anomaly detection based on ZSL and knowledge distillation [316]. Therefore, it is expected that anomaly-detection models will use more ZSL methods in the future to improve the accuracy and generalization ability of GAN-based representation learning.

Table 11
Pixel-level anomaly segmentation AUC ranking for MVTec AD dataset.

Methods	Mean AUC	Pre-trained
CFLOW-AD [297]	0.986	√
Semi-orthogonal [309]	0.982	√
PatchCore [296]	0.981	√
PaDiM [299]	0.975	√
DRAEM [298]	0.973	√
AnoSeg [298]	0.970	×
STPM [303]	0.970	√
InTra [301]	0.969	×
SPADE [308]	0.965	√
NSA [300]	0.963	×
CutPaste [62]	0.960	×
FCDD(semi-supervised) [38]	0.960	√
Patch-SVDD [305]	0.957	×
DFR [156]	0.955	√
IGD [290]	0.930	×
CAVGA(weakly-supervised) [94]	0.930	√
FCDD(unsupervised) [38]	0.920	×
RIAD [306]	0.920	×
UTAD [293]	0.900	×
CAVGA [94]	0.890	×
AE-SSIM [154]	0.870	×

5.2.2. Explainable learning

The research on explainable/interpretable learning aims to make the training process and learning results of networks theoretically interpretable. However, this has been a difficult problem since deep learning was proposed. The inexplicable process of GAN learning, from random noise to the representation of training samples, is a typical black-box problem. Additionally, in anomaly-detection tasks, industrial and infrastructural applications need to be interpretable to match safety requirements. It is also extremely important to medical image analysis to allow these systems to support the decisions of human experts [317].

With the application of deep learning, current research frequently uses informative heatmaps [40], feature vectors [318], individual channels [319], and simplified network structure [38] to explain the decisions of deep neural networks, and these same strategies have been used to try to explain GAN. For example, Bau et al. [37] applied a segmentation-based network dissection method to interfere with the output of the model to the explainable units in the image feature, to better explain the GAN representation units corresponding to a given object. Previous research [320] has demonstrated that interpretable control of GANs can

be realized using a layer-wise image representation, such as controlling over the pose and shape of objects.

The interpretability study of GAN-based anomaly detection has the following purposes:

- Understanding the meaning of the representations that the model learns from the samples and controlling the learning process in a more purposeful manner.
- Understanding the type, location, and severity of anomalies, to better assist human decision-making.
- Quantifying the robustness and credibility of the model and revealing the deficiencies of the current training data.

At present, most explainable anomaly-detection methods belong to the result-oriented interpretation methods, for example, as described in detail in Section 3, using the anomaly score to measure the pixel-wise differences and using explanation heatmaps and activation maps of the receptive field based on pixels to explain the location of the anomaly. The research and application of the interpretability of GAN has been mostly focused on generating content control [321–323] and data enhancement [324]. Thus, proposing an explainable GAN for anomaly prediction is still one of the key directions for future research.

5.2.3. Life-long learning

As a data-driven anomaly-detection method, GAN faces the problem of data imbalance between positive and negative samples, and this problem will remain over time. Abnormal and novel samples do not appear in large numbers until the model finishes its training and is put to use. It is undoubtedly a significant waste of data to ignore the learning of OOD knowledge. Considering that the inductive bias of deep neural networks depends heavily on the distribution of data, the model must have a life-long learning ability to continuously improve the reliability of its anomaly detection.

Life-long learning has the following advantages compared with other methods:

- Reduces the requirements of data collection and annotation in the initial training stage.
- Allows models to be adjusted and extended for more complex tasks while preserving the original knowledge and capabilities.
- Through OOD knowledge, life-long learning can not only improve old knowledge, but also store new knowledge for future use.

However, the learning of OOD knowledge requires adjustments to the model and learning algorithm, which is time-consuming. Meanwhile, the model should not need to be retrained, but only increases the representation learning of new samples as incremental learning [325]. Therefore, methods to maintain efficient life-long learning in the face of unknown abnormalities is an important research direction.

5.2.4. Information bottleneck

As a strong candidate for deep learning interpretation tools, the information bottleneck (IB) method [326] has received extensive attention. After years of development, IB has been used to enhance domain generalization capability [327] and improve the performance of GAN, reinforcement learning, and other tasks [328]. However, at present, only a few anomaly-detection methods refer to the IB method. For example, MemAE regards the encoder as its IB to extract the effective features of samples, thus indicating the feasibility of its model design [134]. Most anomaly-detection methods do not use the IB method to analyze their data, let alone guide the model design by the IB.

IB has the following potential improvements for anomaly detection:

- Provides a more powerful theoretical analysis tool for GAN-based anomaly detection than ALI and self-supervised learning.
- Provides a strong constraint on the latent space of GAN and helps eliminate the deviation between the model distribution and the decision boundary.
- Provides a unified perspective for deep learning and feature engineering to guide the design of networks.

In recent research, DisenIB [143] has been proven to be useful for OOD-detection tasks. This is just the beginning of IB development in the field of anomaly detection. Its development and popularization should capture the attention of researchers.

5.2.5. Vector quantised technology

With the cross fusion of natural language processing (NLP) technology and computer vision, scholars have gradually begun to pay more attention to autoregressive models in image processing. VQ-VAE (Vector Quantized - Variational AutoEncoder) [329] established vector-quantized codebook for encoding in latent space and achieved good image-generation effects using an autoregressive model while simplifying the complexity of image encoding and decoding. This vector-quantized (VQ) technology has a similar implementation in the field of anomaly detection such as MemAE [134] and MNAD [59]. Both implement codebook-like functions through memory modules, but neither is as concise as VQ-VAE.

The use of vector-quantized technology has the following prospects:

- Discrete representation provides the possibility for the use of discrete mathematical tools, which is conducive to more flexible constraints on latent space in the training process.
- Discrete representation reduces the complexity of latent space and facilitates the establishment of more concise model distributions.
- Discrete representation makes the semantic analysis, combination, disassembly, and control more flexible, and is conducive to the measurement, location, and classification of anomalies.

In recent years, VQ technology has made notable progress, for example, high-resolution image generation combined with GAN and Transformer, VQGAN [330], and image generation based on large-scale pre-training model, DALL.E [331]. However, there are no corresponding research achievements in the field of anomaly detection. Therefore, VQ technology is a worthy research direction for anomaly detection.

5.2.6. Transformer

Transformer is currently the most powerful candidate to unify NLP and computer vision [332,333]; thus, it has attracted the attention of many scholars. Transformer has the potential to demonstrate the following advantages in the domain of anomaly detection:

- Common modeling capabilities can be easily adapted to different anomaly-detection tasks.
- Stronger self-attention boosts stronger anomaly-detection performance.
- The inductive bias of Transformer is closer to the human prior and is thus conducive to the establishment of model distributions that are more in line with human needs.

At present, anomaly-detection methods based on Transformer are still scarce, and of them only InTra [301] shows good perfor-

mance. To introduce Transformer into the field of anomaly detection more smoothly, it is necessary to further study its self-supervised representational learning ability, such as a stronger visual tokenizer design, a more appropriate mask-prediction pretext task, and more effective high-level semantic learning strategy. Transformer should be a popular research topic in the future.

6. Conclusions

This review investigated and summarized the research on GAN-based anomaly-detection methods. In this review, we summarized four challenges of deep anomaly detection by reconsidering anomaly, providing a clear evolution of the theories and technologies associated with GAN-based anomaly detection, and discussed the current implementation schemes for anomaly detection and location. The research and applications of GAN anomaly-detection methods in industry, infrastructure, and medical disease were introduced and described in detail. Based on the outstanding issues in the development of GAN-based anomaly detection, we also provided future research directions. We hope that this review will help researchers to develop a comprehensive understanding of GAN as it can be applied to anomaly detection. As more innovative theories and effective techniques are created, we believe that GAN-based anomaly detection will evolve into new forms.

CRediT authorship contribution statement

Xuan Xia: Conceptualization, Writing - original draft, Visualization, Writing - review & editing. **Xizhou Pan:** Writing - original draft, Visualization, Writing - review & editing. **Nan Li:** Visualization, Data curation, Writing - review & editing. **Xing He:** Visualization, Writing - review & editing. **Lin Ma:** Supervision, Validation. **Xiaoguang Zhang:** Data curation, Writing - review & editing. **Ning Ding:** Supervision.

Declaration of Competing Interest

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

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (61806190, 61806191), National Key R&D Program of China (2019YFB1310403), Shenzhen Science and Technology Innovation Council (JCYJ20170410171923840), National Natural Science Foundation of China (U1613227, U1813216), The Guangdong Basic and Applied Basic Research Foundation under Grant (2019A1515111119), and Foundation of Shenzhen Institute of Artificial Intelligence and Robotics for Society (AC01202101022).

References

[1] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM Computing Surveys (CSUR) 41 (2009) 1–58.
[2] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (1998) 2278–2324.
[3] A. Graves, Generating sequences with recurrent neural networks, arXiv preprint arXiv:1308.0850, 2013.
[4] G.E. Hinton, A. Krizhevsky, S.D. Wang, Transforming auto-encoders, in: International conference on artificial neural networks, Springer, 2011, pp. 44–51..
[5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, MIT Press, 2014, pp. 2672–2680.
[6] K. Oksuz, B.C. Cam, S. Kalkan, E. Akbas, Imbalance problems in object detection: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).
[7] D.P. Kingma, M. Welling, Auto-encoding variational bayes, in: International Conference on Learning Representations, 2013..
[8] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville, Improved training of wasserstein gans, arXiv preprint arXiv:1704.00028 (2017)..
[9] Y. Hong, U. Hwang, J. Yoo, S. Yoon, How generative adversarial networks and their variants work: An overview, ACM Computing Surveys (CSUR) 52 (2019) 1–43.
[10] T.W. Cenggoro, Deep learning for imbalance data classification using class expert generative adversarial network, Procedia Computer Science 135 (2018) 60–67.
[11] S. Bulusu, B. Kailkhura, B. Li, P.K. Varshney, D. Song, Anomalous example detection in deep learning: A survey, IEEE Access 8 (2020) 132330–132347.
[12] R. Chalapathy, S. Chawla, Deep learning for anomaly detection: A survey, arXiv preprint arXiv:1901.03407, 2019.
[13] G. Pang, C. Shen, L. Cao, A.V.D. Hengel, Deep learning for anomaly detection: A review, ACM Computing Surveys (CSUR) 54 (2021) 1–38.
[14] F. Di Mattia, P. Galeone, M. De Simoni, E. Ghelfi, A survey on gans for anomaly detection, 2019, arXiv preprint arXiv:1906.11632.
[15] V. Chandola, A. Banerjee, V. Kumar, Outlier detection: A survey, ACM Computing Surveys (CSUR) 14 (2007) 15.
[16] E. Zisselman, A. Tamar, Deep residual flow for out of distribution detection, in, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13994–14003.
[17] M.A. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, Signal Processing 99 (2014) 215–249.
[18] A. Ayadi, O. Ghorbel, A.M. Obeid, M. Abid, Outlier detection approaches for wireless sensor networks: A survey, Computer Networks 129 (2017) 319–333.
[19] C.C. Aggarwal, An Introduction to Outlier Analysis, Springer, 2017, pp. 1–34.
[20] M. Goldstein, S. Uchida, A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data, PloS One 11 (2016) e0152173.
[21] L. Ruff, J.R. Kauffmann, R.A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T.G. Dietterich, K.-R. Müller, A unifying review of deep and shallow anomaly detection, Proceedings of the IEEE (2021).
[22] A. Taha, A.S. Hadi, Anomaly detection methods for categorical data: A review, ACM Computing Surveys (CSUR) 52 (2019) 1–35.
[23] H. Wang, M.J. Bah, M. Hammad, Progress in outlier detection techniques: A survey, IEEE Access 7 (2019) 107964–108000.
[24] R. Foorthuis, On the nature and types of anomalies: A review of deviations in data, 2020, arXiv preprint arXiv:2007.15634.
[25] A. Boukerche, L. Zheng, O. Alfandi, Outlier detection: Methods, models, and classification, ACM Computing Surveys (CSUR) 53 (2020) 1–37.
[26] D. Miljković, Review of novelty detection methods, in: The 33rd International Convention MIPRO, IEEE, 2010, pp. 593–598..
[27] A. Carreño, I. Inza, J.A. Lozano, Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework, Artificial Intelligence Review 53 (2020) 3575–3594.
[28] R. Domingues, M. Filippone, P. Michiardi, J. Zouaoui, A comparative evaluation of outlier detection algorithms: Experiments and analyses, Pattern Recognition 74 (2018) 406–421.
[29] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, Y. Song, Metagan: An adversarial approach to few-shot learning, in: Proceedings of the International Conference on Neural Information Processing Systems, volume 2, 2018, p. 8..
[30] A. Mishra, S. Krishna Reddy, A. Mittal, H.A. Murthy, A generative model for zero shot learning using conditional variational autoencoders, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 2188–2196.
[31] M.H. Bhuyan, D.K. Bhattacharyya, J.K. Kalita, Survey on incremental approaches for network anomaly detection, arXiv preprint arXiv:1211.4493 (2012)..
[32] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, Machine learning: An artificial intelligence approach, Springer Science & Business Media, 2013.
[33] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, T. Aila, Training generative adversarial networks with limited data, arXiv preprint arXiv:2006.06676 (2020)..
[34] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, S. Han, Differentiable augmentation for data-efficient gan training, in: Proceedings of the International Conference on Neural Information Processing Systems, 2020.
[35] I. Golan, R. El-Yaniv, Deep anomaly detection using geometric transformations, in: Proceedings of the International Conference on Neural Information Processing Systems, 2018.
[36] D. Hendrycks, M. Mazeika, S. Kadavath, D. Song, Using self-supervised learning can improve model robustness and uncertainty, in: Proceedings of the International Conference on Neural Information Processing Systems, 2019.
[37] D. Bau, J.-Y. Zhu, H. Strobelt, B. Zhou, J.B. Tenenbaum, W.T. Freeman, A. Torralba, Gan dissection: Visualizing and understanding generative adversarial networks, in: International Conference on Learning Representations, 2019..
[38] P. Liznerski, L. Ruff, R.A. Vandermeulen, B.J. Franks, M. Kloft, K.-R. Müller, Explainable deep one-class classification, in: International Conference on Learning Representations, 2021..
[39] W. Samek, T. Wiegand, K.-R. Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, arXiv preprint arXiv:1708.08296 (2017)..

[40] B. Zhou, Y. Sun, D. Bau, A. Torralba, Interpretable basis decomposition for visual explanation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 119–134.

[41] M. Du, Z. Chen, C. Liu, R. Oak, D. Song, Lifelong anomaly detection through unlearning, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 2019, pp. 1283–1297.

[42] X. Ma, B. Li, Y. Wang, S.M. Erfani, S. Wijewickrema, G. Schoenebeck, D. Song, M.E. Houle, J. Bailey, Characterizing adversarial subspaces using local intrinsic dimensionality, in: International Conference on Learning Representations, 2018..

[43] K. Lee, K. Lee, H. Lee, J. Shin, A simple unified framework for detecting out-of-distribution samples and adversarial attacks, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018.

[44] V. Jumutc, J.A. Suykens, Multi-class supervised novelty detection, IEEE transactions on pattern analysis and machine intelligence 36 (2014) 2510–2523.

[45] R. Feinman, R.R. Curtin, S. Shintre, A.B. Gardner, Detecting adversarial samples from artifacts, in: International Conference on Machine Learning, 2017..

[46] S. Kim, Y. Choi, M. Lee, Deep learning with support vector data description, Neurocomputing 165 (2015) 111–117.

[47] H. Song, Z. Jiang, A. Men, B. Yang, A hybrid semi-supervised anomaly detection model for high-dimensional data, Computational intelligence and neuroscience 2017 (2017).

[48] H. Wu, S. Prasad, Semi-supervised deep learning using pseudo labels for hyperspectral image classification, IEEE Transactions on Image Processing 27 (2017) 1259–1270.

[49] P. Perera, V.M. Patel, Learning deep features for one-class classification, IEEE Transactions on Image Processing 28 (2019) 5450–5463.

[50] J. Liu, K. Song, M. Feng, Y. Yan, Z. Tu, L. Zhu, Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection, Optics and Lasers in Engineering 136 (2021) 106324.

[51] F. Gao, J. Li, R. Cheng, Y. Zhou, Y. Ye, Connet: Deep semi-supervised anomaly detection based on sparse positive samples, IEEE Access 9 (2021) 67249–67258.

[52] D. Hendrycks, K. Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, in: International Conference on Learning Representations, 2017..

[53] M. Sakurada, T. Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, 2014, pp. 4–11.

[54] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, Special Lecture on IE 2 (2015) 1–18.

[55] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, Long short term memory networks for anomaly detection in time series, in: 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, volume 89, Presses universitaires de Louvain, 2015, pp. 89–94..

[56] T. Schlegl, P. Seeböck, S.M. Waldstein, U. Schmidt-Erfurth, G. Langs, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: International conference on information processing in medical imaging, Springer, 2017, pp. 146–157..

[57] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S.A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: D. Jennifer, K. Andreas (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 PMLR, 2018, pp. 4393–4402, URL:http://proceedings.mlr.press.

[58] D. Abati, A. Porrello, S. Calderara, R. Cucchiara, Latent space autoregression for novelty detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 481–490.

[59] H. Park, J. Noh, B. Ham, Learning memory-guided normality for anomaly detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14372–14381.

[60] A. Markovitz, G. Sharir, I. Friedman, L. Zelnik-Manor, S. Avidan, Graph embedded pose clustering for anomaly detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10539–10547.

[61] S. Goyal, A. Raghunathan, M. Jain, H.V. Simhadri, P. Jain, Drocc: Deep robust one-class classification, in: International Conference on Machine Learning, PMLR, 2020, pp. 3711–3721.

[62] C.-L. Li, K. Sohn, J. Yoon, T. Pfister, Cutpaste: Self-supervised learning for anomaly detection and localization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.

[63] M. Rudolph, T. Wehrbein, B. Rosenhahn, B. Wandt, Fully convolutional cross-scale-flows for image-based defect detection, arXiv preprint arXiv:2110.02855 (2021)..

[64] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: International Conference on Learning Representations, 2015..

[65] M. Mirza, S. Osindero, Conditional generative adversarial nets, Computer Science (2014) 2672–2680.

[66] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016, pp. 2180–2188.

[67] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: International conference on machine learning, PMLR, 2017, pp. 214–223..

[68] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[69] E. Kodirov, T. Xiang, S. Gong, Semantic autoencoder for zero-shot learning, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3174–3183.

[70] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[71] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.

[72] J. Donahue, P. Krähenbühl, T. Darrell, Adversarial feature learning, in: International Conference on Learning Representations, 2017..

[73] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, A. Courville, Adversarially learned inference, in: International Conference on Learning Representations, 2017..

[74] Q. Xie, Z. Dai, Y. Du, E. Hovy, G. Neubig, Controllable invariance through adversarial feature learning, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2017.

[75] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, N. Sebe, Abnormal event detection in videos using generative adversarial nets, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 1577–1581.

[76] M. Ravanbakhsh, E. Sangineto, M. Nabi, N. Sebe, Training adversarial discriminators for cross-channel abnormal event detection in crowds, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019, pp. 1896–1904.

[77] M. Sabokrou, M. Khalooei, M. Fathy, E. Adeli, Adversarially learned one-class classifier for novelty detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3379–3388.

[78] H.-G. Wang, X. Li, T. Zhang, Generative adversarial network based novelty detection using minimized reconstruction error, Frontiers of Information Technology & Electronic Engineering 19 (2018) 116–125.

[79] H. Zenati, C.S. Foo, B. Lecouat, G. Manek, V.R. Chandrasekhar, Efficient gan-based anomaly detection, in: Proceedings of the 20th IEEE International Conference on Data Mining (ICDM), 2018.

[80] S. Akcay, A. Atapour-Abarghouei, T.P. Breckon, Ganomaly: Semi-supervised anomaly detection via adversarial training, in: Asian conference on computer vision, Springer, 2018, pp. 622–637..

[81] C. Li, H. Liu, C. Chen, Y. Pu, L. Chen, R. Henao, L. Carin, Alice: Towards understanding adversarial learning for joint distribution matching, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2017.

[82] H. Zenati, M. Romain, C.-S. Foo, B. Lecouat, V. Chandrasekhar, Adversarially learned anomaly detection, in: 2018 IEEE International conference on data mining (ICDM), IEEE, 2018, pp. 727–736.

[83] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, in: International Conference on Learning Representations, 2018..

[84] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial networks, in: International conference on machine learning, PMLR, 2019, pp. 7354–7363..

[85] I. Haloui, J.S. Gupta, V. Feuillard, Anomaly detection with wasserstein gan, arXiv preprint arXiv:1812.02463 (2018)..

[86] T. Schlegl, P. Seeböck, S.M. Waldstein, G. Langs, U. Schmidt-Erfurth, f-anogan: Fast unsupervised anomaly detection with generative adversarial networks, Medical image analysis 54 (2019) 30–44.

[87] M. Ducoffe, I. Haloui, J.S. Gupta, I. Supaero, Anomaly detection on time series with wasserstein gan applied to phm, International Journal of Prognostics and Health Management (2019).

[88] C. Zhong, K. Yan, Y. Dai, N. Jin, B. Lou, Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks, Energies 12 (2019) 527.

[89] C. Chen, P. Chen, H. Song, Y. Tao, Y. Xie, S. Ding, L. Ma, Anomaly detection by one class latent regularized networks, 2020, arXiv preprint arXiv:2002.01607.

[90] S. Pidhorskyi, R. Almohsen, D.A. Adjeroh, G. Doretto, Generative probabilistic novelty detection with adversarial autoencoders, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018.

[91] K. Zhou, S. Gao, J. Cheng, Z. Gu, H. Fu, Z. Tu, J. Yang, Y. Zhao, J. Liu, Sparse-gan: Sparsity-constrained generative adversarial network for anomaly detection in retinal oct image, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1227–1231.

[92] P. Perera, R. Nallapati, B. Xiang, Ocgan: One-class novelty detection using gans with constrained latent representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2898–2906.

[93] D. Kimura, S. Chaudhury, M. Narita, A. Munawar, R. Tachibana, Adversarial discriminative attention for robust anomaly detection, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 2172–2181.

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

[95] M.I. Belghazi, S. Rajeswar, O. Mastropietro, N. Rostamzadeh, J. Mitrovic, A. Courville, Hierarchical adversarially learned inference, in: International Conference on Learning Representations, 2018..

[96] M. Rosca, B. Lakshminarayanan, S. Mohamed, Distribution matching in variational inference, arXiv preprint arXiv:1802.06847 (2018)..
[97] J. Donahue, K. Simonyan, Large scale adversarial representation learning, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2019.
[98] A.H. Li, Y. Wang, C. Chen, J. Gao, Decomposed adversarial learned inference, arXiv preprint arXiv:2004.10267 (2020)..
[99] Y. Dandi, H. Bharadhwaj, A. Kumar, P. Rai, Generalized adversarially learned inference, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021.
[100] D. Hendrycks, M. Mazeika, S. Kadavath, D. Song, Using self-supervised learning can improve model robustness and uncertainty, in: Proceedings of the International Conference on Neural Information Processing Systems, 2019.
[101] P. Perera, V.I. Morariu, R. Jain, V. Manjunatha, C. Wigington, V. Ordonez, V.M. Patel, Generative-discriminative feature representations for open-set recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11814–11823.
[102] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9729–9738.
[103] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International conference on machine learning PMLR, 2020, pp. 1597–1607.
[104] J. Tack, S. Mo, J. Jeong, J. Shin, Csi: Novelty detection via contrastive learning on distributionally shifted instances, in: Proceedings of the International Conference on Neural Information Processing Systems, 2020.
[105] T. Jiang, W. Xie, Y. Li, J. Lei, Q. Du, Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection, IEEE Transactions on Neural Networks and Learning Systems (2021).
[106] Y.-H. Nho, S. Ryu, D.-S. Kwon, Ui-gan: Generative adversarial network-based anomaly detection using user initial information for wearable devices, IEEE Sensors Journal 21 (2021) 9949–9958.
[107] S. Motamed, P. Rogalla, F. Khalvati, Randgan: randomized generative adversarial network for detection of covid-19 in chest x-ray, Scientific Reports 11 (2021) 1–10.
[108] K. Storey-Fisher, M. Huertas-Company, N. Ramachandra, F. Lanusse, A. Leauthaud, Y. Luo, S. Huang, J.X. Prochaska, Anomaly detection in hyper suprime-cam galaxy images with generative adversarial networks, arXiv preprint arXiv:2105.02434 (2021)..
[109] Y. Chen, Q. Gao, X. Wang, Inferential wasserstein generative adversarial networks, arXiv preprint arXiv:2109.05652 (2021)..
[110] Z. Yue, T. Wang, H. Zhang, Q. Sun, X.-S. Hua, Counterfactual zero-shot and open-set visual recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[111] S. Kong, D. Ramanan, Opengan: Open-set recognition via open data generation, arXiv preprint arXiv:2104.02939 (2021)..
[112] T. Kim, M. Cha, H. Kim, J.K. Lee, J. Kim, Learning to discover cross-domain relations with generative adversarial networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 1857–1865.
[113] Z. Yi, H. Zhang, P. Tan, M. Gong, Dualgan: Unsupervised dual learning for image-to-image translation, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2849–2857.
[114] R. Zhang, P. Isola, A.A. Efros, Colorful image colorization, in: European conference on computer vision, Springer, 2016, pp. 649–666..
[115] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.
[116] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A.A. Efros, Context encoders: Feature learning by inpainting, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536–2544.
[117] R. Zhang, P. Isola, A.A. Efros, Split-brain autoencoders: Unsupervised learning by cross-channel prediction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1058–1067.
[118] N. Komodakis, S. Gidaris, Unsupervised representation learning by predicting image rotations, in: International Conference on Learning Representations, 2018..
[119] M. Noroozi, P. Favaro, Unsupervised learning of visual representations by solving jigsaw puzzles, in: European conference on computer vision, Springer, 2016, pp. 69–84.
[120] C. Doersch, A. Gupta, A.A. Efros, Unsupervised visual representation learning by context prediction, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1422–1430.
[121] T. Chen, X. Zhai, M. Ritter, M. Lucic, N. Houlsby, Self-supervised gans via auxiliary rotation loss, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 12154–12163.
[122] R. Ali, M.U.K. Khan, C.M. Kyung, Self-supervised representation learning for visual anomaly detection, arXiv preprint arXiv:2006.09654 (2020)..
[123] H. Nakanishi, M. Suzuki, Y. Matsuo, Iterative image inpainting with structural similarity mask for anomaly detection, in: International Conference on Learning Representations, 2021..
[124] N.-T. Tran, V.-H. Tran, N.-B. Nguyen, T.-K. Nguyen, N.-M. Cheung, On data augmentation for gan training, IEEE Transactions on Image Processing 30 (2021) 15.
[125] L. Bergman, Y. Hoshen, Classification-based anomaly detection for general data, in: International Conference on Learning Representations, 2020..
[126] Y. Fei, C. Huang, C. Jinkun, M. Li, Y. Zhang, C. Lu, Attribute restoration framework for anomaly detection, IEEE Transactions on Multimedia (2020).
[127] P. Perera, V. Patel, A joint representation learning and feature modeling approach for one-class recognition, in: International Conference on Pattern Recognition, 2021..
[128] P. Oza, H.V. Nguyen, V.M. Patel, Multiple class novelty detection under data distribution shift, in: European Conference on Computer Vision, Springer, 2020, pp. 432–449.
[129] Z. Zhao, B. Li, R. Dong, P. Zhao, A surface defect detection method based on positive samples, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2018, pp. 473–481..
[130] M. Salehi, A. Eftekhar, N. Sadjadi, M.H. Rohban, H.R. Rabiee, Puzzle-ae: Novelty detection in images through solving puzzles, arXiv preprint arXiv:2008.12959 (2020)..
[131] E. Wong, L. Rice, J.Z. Kolter, Fast is better than free: Revisiting adversarial training, arXiv preprint arXiv:2001.03994 (2020)..
[132] J. Song, K. Kong, Y.-I. Park, S.-G. Kim, S.-J. Kang, Anoseg: Anomaly segmentation network using self-supervised learning, arXiv preprint arXiv:2110.03396 (2021)..
[133] X. Chen, K. He, Exploring simple siamese representation learning, arXiv preprint arXiv:2011.10566 (2020)..
[134] D. Gong, L. Liu, V. Le, B. Saha, M.R. Mansour, S. Venkatesh, A.v.d. Hengel, Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1705–1714.
[135] K. Komoto, H. Aizawa, K. Kato, Consistency ensured bi-directional gan for anomaly detection, in: International Workshop on Frontiers of Computer Vision, Springer, 2020, pp. 236–247..
[136] R. La Grassa, I. Gallo, N. Landro, Ocmst: One-class novelty detection using convolutional neural network and minimum spanning trees, arXiv preprint arXiv:2003.13524 (2020)..
[137] M.A.N. Oz, M. Mercimek, O.T. Kaymakci, Anomaly localization in regular textures based on deep convolutional generative adversarial networks, Applied Intelligence (2021) 1–10.
[138] M.Z. Zaheer, J.-H. Lee, M. Astrid, S.-I. Lee, Old is gold: Redefining the adversarially learned one-class classifier training paradigm, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14183–14193.
[139] Z. Zhang, S. Chen, L. Sun, P-kdgan: Progressive knowledge distillation with gans for one-class novelty detection, in: International Joint Conference on Artificial Intelligence, 2020..
[140] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S.A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: International conference on machine learning PMLR, 2018, pp. 4393–4402.
[141] Y.-C. Hsu, Y. Shen, H. Jin, Z. Kira, Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10951–10960.
[142] S. Liang, Y. Li, R. Srikant, Enhancing the reliability of out-of-distribution image detection in neural networks, arXiv preprint arXiv:1706.02690 (2017)..
[143] Z. Pan, L. Niu, J. Zhang, L. Zhang, Disentangled information bottleneck, arXiv preprint arXiv:2012.07372 (2020)..
[144] G. Shalev, Y. Adi, J. Keshet, Out-of-distribution detection using multiple semantic label representations, arXiv preprint arXiv:1808.06664 (2018)..
[145] A. Vyas, N. Jammalamadaka, X. Zhu, D. Das, B. Kaul, T.L. Willke, Out-of-distribution detection using an ensemble of self supervised leave-out classifiers, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 550–564.
[146] J. Winkens, R. Bunel, A.G. Roy, R. Stanforth, V. Natarajan, J.R. Ledsam, P. MacWilliams, P. Kohli, A. Karthikesalingam, S. Kohl, Contrastive training for improved out-of-distribution detection, arXiv preprint arXiv:2007.05566 (2020)..
[147] Q. Yu, K. Aizawa, Unsupervised out-of-distribution detection by maximum classifier discrepancy, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9518–9526.
[148] S. Akçay, A. Atapour-Abarghouei, T.P. Breckon, Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
[149] Z. Chen, J. Duan, L. Kang, G. Qiu, Supervised anomaly detection via conditional generative adversarial network and ensemble active learning, arXiv preprint arXiv:2104.11952 (2021)..
[150] J. Wang, G. Yi, S. Zhang, Y. Wang, An unsupervised generative adversarial network-based method for defect inspection of texture surfaces, Applied Sciences 11 (2021) 283.
[151] C.X. Ling, J. Huang, H. Zhang, Auc: a statistically consistent and more discriminating measure than accuracy, Ijcai 3 (2003) 519–524.
[152] Z.-C. Qin, Roc analysis for predictions made by probabilistic classifiers, in: 2005 International Conference on Machine Learning and Cybernetics, volume 5, IEEE, 2005, pp. 3119–3124.

[153] D. Dehaene, O. Frigo, S. Combrexelle, P. Eline, Iterative energy-based projection on a normal data manifold for anomaly localization, arXiv preprint arXiv:2002.03734 (2020)..
[154] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, Mvtec ad - a comprehensive real-world dataset for unsupervised anomaly detection, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[155] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4183–4192.
[156] Y. Shi, J. Yang, Z. Qi, Unsupervised anomaly segmentation via deep feature reconstruction, Neurocomputing 424 (2021) 9–22.
[157] F. Carrara, G. Amato, L. Brombin, F. Falchi, C. Gennaro, Combining gans and autoencoders for efficient anomaly detection, in: 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 3939–3946.
[158] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.
[159] M. Hamghalam, B. Lei, T. Wang, High tissue contrast mri synthesis using multi-stage attention-gan for glioma segmentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020.
[160] C. Qi, J. Chen, G. Xu, Z. Xu, T. Lukasiewicz, Y. Liu, Sag-gan: Semi-supervised attention-guided gans for data augmentation on medical images, arXiv preprint arXiv:2011.07534 (2020)..
[161] 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 20 (2020) 1459.
[162] Q. Luo, X. Fang, L. Liu, C. Yang, Y. Sun, Automated visual defect detection for flat steel surface: A survey, IEEE Transactions on Instrumentation and Measurement 69 (2020) 626–644.
[163] K. Liu, A. Li, X. Wen, H. Chen, P. Yang, Steel surface defect detection using gan and one-class classifier, in: 2019 25th International Conference on Automation and Computing (ICAC), IEEE, 2019, pp. 1–6.
[164] Y. Lai, J.-S. Hu, Y. Tsai, W. Chiu, Industrial anomaly detection and one-class classification using generative adversarial networks, in: 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), IEEE, 2018, pp. 1444–1449.
[165] L. Liu, D. Cao, Y. Wu, T. Wei, Defective samples simulation through adversarial training for automatic surface inspection, Neurocomputing 360 (2019) 230–245.
[166] S. Niu, B. Li, X. Wang, H. Lin, Defect image sample generation with gan for improving defect recognition, IEEE Transactions on Automation Science and Engineering 17 (2020) 1611–1622.
[167] H. Di, X. Ke, Z. Peng, Z. Dongdong, Surface defect classification of steels with a new semi-supervised learning method, Optics and Lasers in Engineering 117 (2019) 40–48.
[168] R. Skilton, Y. Gao, Visual detection of generic defects in industrial components using generative adversarial networks, in: 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), IEEE, 2019, pp. 489–494.
[169] R. Guo, H. Liu, G. Xie, Y. Zhang, Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning, IEEE Sensors Journal 21 (2021) 10844–10853.
[170] J. Balzategui, L. Eciolaza, D. Maestro-Watson, Anomaly detection and automatic labeling for solar cell quality inspection based on generative adversarial network, arXiv preprint arXiv:2103.03518 (2021)..
[171] S. Song, K. Yang, A. Wang, S. Zhang, M. Xia, A mura detection model based on unsupervised adversarial learning, IEEE Access 9 (2021) 49920–49928.
[172] C. Xie, K. Yang, A. Wang, C. Chen, W. Li, A mura detection method based on an improved generative adversarial network, IEEE Access 9 (2021) 68826–68836.
[173] Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen, Automatic road crack detection using random structured forests, IEEE Transactions on Intelligent Transportation Systems 17 (2016) 3434–3445.
[174] 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 34 (2015) 1–12.
[175] K. Song, Y. Yan, Micro surface defect detection method for silicon steel strip based on saliency convex active contour model, Mathematical Problems in Engineering 2013 (2013).
[176] P. Xu, R. Du, Z. Zhang, Predicting pipeline leakage in petrochemical system through gan and lstm, Knowledge-Based Systems 175 (2019) 50–61.
[177] S. Cao, L. Wen, X. Li, L. Gao, Application of generative adversarial networks for intelligent fault diagnosis, in: 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), IEEE, 2018, pp. 711–715.
[178] O. Silvén, M. Niskanen, H. Kauppinen, Wood inspection with non-supervised clustering, Machine Vision and Applications 13 (2003) 275–285.
[179] M. Wieler, T. Hahn, Weakly supervised learning for industrial optical inspection, in: DAGM symposium, 2007.
[180] K. Song, Y. Yan, A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects, Applied Surface Science 285 (2013) 858–864.
[181] K.-C. Song, S.-P. Hu, Y.-H. Yan, J. Li, Surface defect detection method using saliency linear scanning morphology for silicon steel strip under oil pollution interference, Isij International 54 (2014) 2598–2607.
[182] J. Gan, Q. Li, J. Wang, H. Yu, A hierarchical extractor-based visual rail surface inspection system, IEEE Sensors Journal 17 (2017) 7935–7944.
[183] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9592–9600.
[184] D. Tabernik, S. Šela, J. Skvarč, D. Skočaj, Segmentation-based deep-learning approach for surface-defect detection, Journal of Intelligent Manufacturing 31 (2020) 759–776.
[185] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mechanical Systems and Signal Processing 64 (2015) 100–131.
[186] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido, T. Nakamura, Y. Kawaguchi, Mimii due: Sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions, in: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2021.
[187] X. Zhou, J. Xiong, X. Zhang, X. Liu, J. Wei, A radio anomaly detection algorithm based on modified generative adversarial network, IEEE Wireless Communications Letters (2021).
[188] C. Li, D. Cabrera, F. Sancho, R.-V. Sánchez, M. Cerrada, J. Long, J.V. de Oliveira, Fusing convolutional generative adversarial encoders for 3d printer fault detection with only normal condition signals, Mechanical Systems and Signal Processing 147 (2021) 107108.
[189] K. Yan, Chiller fault detection and diagnosis with anomaly detective generative adversarial network, Building and Environment 107982 (2021).
[190] B. Li, F. Cheng, H. Cai, X. Zhang, W. Cai, A semi-supervised approach to fault detection and diagnosis for building hvac systems based on the modified generative adversarial network, Energy and Buildings 246 (2021) 111044.
[191] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Advanced Engineering Informatics 29 (2015) 196–210.
[192] A. Mohan, S. Poobal, Crack detection using image processing: A critical review and analysis, Alexandria Engineering Journal 57 (2018) 787–798.
[193] L. Zhang, F. Yang, Y.D. Zhang, Y.J. Zhu, Road crack detection using deep convolutional neural network, in: 2016 IEEE international conference on image processing (ICIP), IEEE, 2016, pp. 3708–3712.
[194] W. Cao, Q. Liu, Z. He, Review of pavement defect detection methods, IEEE Access 8 (2020) 14531–14544.
[195] W. Wang, M. Wang, H. Li, H. Zhao, K. Wang, C. He, J. Wang, S. Zheng, J. Chen, Pavement crack image acquisition methods and crack extraction algorithms: A review, Journal of Traffic and Transportation Engineering (English Edition) 6 (2019) 535–556.
[196] Q. Mei, M. Gül, A cost effective solution for pavement crack inspection using cameras and deep neural networks, Construction and Building Materials 256 (2020) 119397.
[197] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
[198] K. Zhang, Y. Zhang, H. Cheng, Self-supervised structure learning for crack detection based on cycle-consistent generative adversarial networks, Journal of Computing in Civil Engineering 34 (2020) 04020004.
[199] K. Zhang, Y. Zhang, H.-D. Cheng, Crackgan: Pavement crack detection using partially accurate ground truths based on generative adversarial learning, IEEE Transactions on Intelligent Transportation Systems (2020).
[200] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241..
[201] W. Zhai, J. Zhu, Y. Cao, Z. Wang, A generative adversarial network based framework for unsupervised visual surface inspection, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 1283–1287.
[202] H. Oliveira, P.L. Correia, Crackit–an image processing toolbox for crack detection and characterization, in: 2014 IEEE international conference on image processing (ICIP), IEEE, 2014, pp. 798–802.
[203] Z. Gao, B. Peng, T. Li, C. Gou, Generative adversarial networks for road crack image segmentation, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–8.
[204] S. Chambon, J. Moliard, Automatic road pavement assessment with image processing: Review and comparison, International Journal of Geophysics (2011).
[205] L. Duan, H. Geng, J. Pang, J. Zeng, Unsupervised pixel-level crack detection based on generative adversarial network, in: Proceedings of the 2020 5th International Conference on Multimedia Systems and Signal Processing, 2020, pp. 6–10.
[206] J. Mao, H. Wang, B.F. Spencer Jr, Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders, Structural Health Monitoring 1475921720924601 (2020).
[207] K. Lee, D.H. Shin, Generative model of acceleration data for deep learning-based damage detection for bridges using generative adversarial network, Journal of KIBIM 9 (2019) 42–51.
[208] R. Amhaz, S. Chambon, J. Idier, V. Baltazart, Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path

selection, IEEE Transactions on Intelligent Transportation Systems 17 (2016) 2718–2729.
[209] L.A. Gatys, A.S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2414–2423.
[210] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, H. Omata, Generative adversarial network for road damage detection, Computer-Aided Civil and Infrastructure Engineering 36 (2021) 47–60.
[211] S. Alahakoon, Y.Q. Sun, M. Spiryagin, C. Cole, Rail flaw detection technologies for safer, reliable transportation: a review, Journal of Dynamic Systems, Measurement, and Control 140 (2018).
[212] S. Liu, Q. Wang, Y. Luo, A review of applications of visual inspection technology based on image processing in the railway industry, Transportation Safety and Environment 1 (2019) 185–204.
[213] P. Yang, W. Jin, P. Tang, Anomaly detection of railway catenary based on deep convolutional generative adversarial networks, in: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), IEEE, 2018, pp. 1366–1370.
[214] Y. Lyu, Z. Han, J. Zhong, C. Li, Z. Liu, A generic anomaly detection of catenary support components based on generative adversarial networks, IEEE Transactions on Instrumentation and Measurement 69 (2019) 2439–2448.
[215] Y. Lyu, Z. Han, J. Zhong, C. Li, Z. Liu, A gan-based anomaly detection method for isoelectric line in high-speed railway, in: 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, 2019, pp. 1–6.
[216] L. Xue, S. Gao, Unsupervised anomaly detection system for railway turnout based on gan, Journal of Physics: Conference Series, volume 1345, IOP Publishing (2019) 032069.
[217] K. Wang, X. Zhang, Q. Hao, Y. Wang, Y. Shen, Application of improved least-square generative adversarial networks for rail crack detection by ae technique, Neurocomputing 332 (2019) 236–248.
[218] V.N. Nguyen, R. Jenssen, D. Roverso, Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning, International Journal of Electrical Power & Energy Systems 99 (2018) 107–120.
[219] R. Jenssen, D. Roverso, Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning, International Journal of Electrical Power & Energy Systems 99 (2018) 107–120.
[220] X. Liu, X. Miao, H. Jiang, J. Chen, Review of data analysis in vision inspection of power lines with an in-depth discussion of deep learning technology, arXiv preprint arXiv:2003.09802 (2020)..
[221] L. Luo, W. Hsu, S. Wang, Data augmentation using generative adversarial networks for electrical insulator anomaly detection, in: Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering, 2020, pp. 231–236.
[222] W. Chang, G. Yang, E. Li, Z. Liang, Toward a cluttered environment for learning-based multi-scale overhead ground wire recognition, Neural Processing Letters 48 (2018) 1789–1800.
[223] W. Chang, G. Yang, J. Yu, Z. Liang, Real-time segmentation of various insulators using generative adversarial networks, IET Computer Vision 12 (2018) 596–602.
[224] W. Chang, G. Yang, Z. Wu, Z. Liang, Learning insulators segmentation from synthetic samples, in: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–7.
[225] I. Siniosoglou, P. Radoglou-Grammatikis, G. Efstathopoulos, P. Fouliras, P. Sarigiannidis, A unified deep learning anomaly detection and classification approach for smart grid environments, IEEE Transactions on Network and Service Management (2021).
[226] L. Zhang, H. Wei, Z. Lyu, H. Wei, P. Li, A small-sample faulty line detection method based on generative adversarial networks, Expert Systems with Applications 169 (2021) 114378.
[227] W. Lawson, E. Bekele, K. Sullivan, Finding anomalies with generative adversarial networks for a patrolbot, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 12–13.
[228] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, M. Chen, Generative adversarial network based abnormal behavior detection in massive crowd videos: a hajj case study, Journal of Ambient Intelligence and Humanized Computing (2021) 1–12.
[229] D. Chen, L. Yue, X. Chang, M. Xu, T. Jia, Nm-gan: Noise-modulated generative adversarial network for video anomaly detection, Pattern Recognition 116 (2021) 107969.
[230] T. Ganokratanaa, S. Aramvith, N. Sebe, Anomaly event detection using generative adversarial network for surveillance videos, in: 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 2019, pp. 1395–1399.
[231] Y. Sun, W. Yu, Y. Chen, A. Kadam, Time series anomaly detection based on gan, in: 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), IEEE, 2019, pp. 375–382..
[232] Y. Qiu, T. Misu, C. Busso, Driving anomaly detection with conditional generative adversarial network using physiological and can-bus data, in: 2019 International Conference on Multimodal Interaction, 2019, pp. 164–173.
[233] Z. Situ, S. Teng, H. Liu, J. Luo, Q. Zhou, Automated sewer defects detection using style-based generative adversarial networks and fine-tuned well-known cnn classifier, IEEE Access 9 (2021) 59498–59507.
[234] J. Zhong, W. Xie, Y. Li, J. Lei, Q. Du, Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection, IEEE Transactions on Geoscience and Remote Sensing (2020).
[235] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, C. Fookes, Deep learning for medical anomaly detection–a survey, arXiv preprint arXiv:2012.02364 (2020)..
[236] J.M. Wolterink, K. Kamnitsas, C. Ledig, I.Išgum, Generative adversarial networks and adversarial methods in biomedical image analysis, arXiv preprint arXiv:1810.10352 (2018)..
[237] X. Yi, E. Walia, P. Babyn, Generative adversarial network in medical imaging: A review, Medical image analysis 58 (2019) 101552.
[238] Y. Xue, T. Xu, H. Zhang, L.R. Long, X. Huang, Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation, Neuroinformatics 16 (2018) 383–392.
[239] L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan, J. Paisley, An adversarial learning approach to medical image synthesis for lesion detection, IEEE journal of biomedical and health informatics 24 (2020) 2303–2314.
[240] J. Wolleb, R. Sandkühler, P.C. Cattin, Descargan: Disease-specific anomaly detection with weak supervision, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 14–24.
[241] J. Song, H. Wang, Y. Liu, W. Wu, G. Dai, Z. Wu, P. Zhu, W. Zhang, K.W. Yeom, K. Deng, End-to-end automatic differentiation of the coronavirus disease 2019 (covid-19) from viral pneumonia based on chest ct, European journal of nuclear medicine and molecular imaging 47 (2020) 2516–2524..
[242] A. Madani, M. Moradi, A. Karargyris, T. Syeda-Mahmood, Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation, in: 2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018), IEEE, 2018, pp. 1038–1042.
[243] Z. Han, B. Wei, A. Mercado, S. Leung, S. Li, Spine-gan: Semantic segmentation of multiple spinal structures, Medical image analysis 50 (2018) 23–35.
[244] A. Işín, C. Direkoğlu, M. Şah, Review of mri-based brain tumor image segmentation using deep learning methods, Procedia Computer Science 102 (2016) 317–324.
[245] V. Alex, M.S. KP, S.S. Chennamsetty, G. Krishnamurthi, Generative adversarial networks for brain lesion detection, in: Medical Imaging 2017: Image Processing, volume 10133, International Society for Optics and Photonics, 2017, p. 101330G..
[246] C. Baur, B. Wiestler, S. Albarqouni, N. Navab, Deep autoencoding models for unsupervised anomaly segmentation in brain mr images, in: International MICCAI Brainlesion Workshop, Springer, 2018, pp. 161–169.
[247] X. Chen, E. Konukoglu, Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders, in: International conference on Medical Imaging with Deep Learning, 2018.
[248] M. Rezaei, K. Harmuth, W. Gierke, T. Kellermeier, M. Fischer, H. Yang, C. Meinel, A conditional adversarial network for semantic segmentation of brain tumor, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 241–252..
[249] C. Han, L. Rundo, K. Murao, Z. Milacski, S. Satoh, Gan-based multiple adjacent brain mri slice reconstruction for unsupervised alzheimer's disease diagnosis, in: International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics, Springer, Cham, 2019, pp. 44–54.
[250] C.F. Baumgartner, L.M. Koch, K.C. Tezcan, J.X. Ang, E. Konukoglu, Visual feature attribution using wasserstein gans, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8309–8319.
[251] S. Lin, F. Qin, Y. Li, R.A. Bly, K.S. Moe, B. Hannaford, Lc-gan: Image-to-image translation based on generative adversarial network for endoscopic images, in: International Conference on Intelligent Robots and Systems, 2020.
[252] A. Swiecicki, N. Konz, M. Buda, M.A. Mazurowski, A generative adversarial network-based abnormality detection using only normal images for model training with application to digital breast tomosynthesis, Scientific reports 11 (2021) 1–13.
[253] M. Loey, F. Smarandache, N.E.M. Khalifa, Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning, Symmetry 12 (2020) 651.
[254] M. Loey, G. Manogaran, N.E.M. Khalifa, A deep transfer learning model with classical data augmentation and cgan to detect covid-19 from chest ct radiography digital images, Neural Computing and Applications (2020) 1–13.
[255] N.E.M. Khalifa, M.H.N. Taha, A.E. Hassanien, S. Elghamrawy, Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset, arXiv preprint arXiv:2004.01184 (2020)..
[256] L. Zhang, B. Shen, A. Barnawi, S. Xi, N. Kumar, Y. Wu, Feddpgan: Federated differentially private generative adversarial networks framework for the detection of covid-19 pneumonia, Information Systems Frontiers (2021) 1–13.
[257] L. Zhang, A. Gooya, A.F. Frangi, Semi-supervised assessment of incomplete lv coverage in cardiac mri using generative adversarial nets, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 61–68.
[258] S. Aida, J. Okugawa, S. Fujisaka, T. Kasai, H. Kameda, T. Sugiyama, Deep learning of cancer stem cell morphology using conditional generative adversarial networks, Biomolecules 10 (2020) 931.

[259] M. Tuba, E. Tuba, Generative adversarial optimization (goa) for acute lymphocytic leukemia detection, Studies in Informatics and Control 28 (2019) 245–254.
[260] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, J.-P. Radtke, K. Maier-Hein, Adversarial networks for the detection of aggressive prostate cancer, in: International conference on medical image computing and computer-assisted intervention, 2017.
[261] A. Udrea, G.D. Mitra, Generative adversarial neural networks for pigmented and non-pigmented skin lesions detection in clinical images, in: 2017 21st International Conference on Control Systems and Computer Science (CSCS), IEEE, 2017, pp. 364–368.
[262] J. Deng, N. Cummins, M. Schmitt, K. Qian, F. Ringeval, B. Schuller, Speech-based diagnosis of autism spectrum condition by generative adversarial network representations, in: Proceedings of the 2017 International Conference on Digital Health, 2017, pp. 53–57.
[263] Y.-H. Nho, J.G. Lim, D.-S. Kwon, Cluster-analysis-based user-adaptive fall detection using fusion of heart rate sensor and accelerometer in a wearable device, IEEE Access 8 (2020) 40389–40401.
[264] G. Vavoulas, M. Pediaditis, E.G. Spanakis, M. Tsiknakis, The mobifall dataset: An initial evaluation of fall detection algorithms using smartphones, in: 13th IEEE International Conference on BioInformatics and BioEngineering, IEEE, 2013, pp. 1–4.
[265] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, M. Tsiknakis, The mobiact dataset: Recognition of activities of daily living using smartphones, in: International Conference on Information and Communication Technologies for Ageing Well and e-Health, volume 2, SciTePress, 2016, pp. 143–151.
[266] V. Morath, M. Keuper, M. Rodriguez-Franco, S. Deswal, G. Fiala, B. Blumenthal, D. Kaschek, J. Timmer, G. Neuhaus, S. Ehl, Semi-automatic determination of cell surface areas used in systems biology, Front. Biosci. (Elite Ed.) 5 (2013) 533–545.
[267] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, A.A. Argyros, Fire: fundus image registration dataset, Modeling and Artificial Intelligence in Ophthalmology 1 (2017) 16–28.
[268] D. Li, D. Chen, J. Goh, S.-K. Ng, Anomaly detection with generative adversarial networks for multivariate time series, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018.
[269] Z. Lin, Y. Shi, Z. Xue, Idsgan: Generative adversarial networks for attack generation against intrusion detection, arXiv preprint arXiv:1809.02077 (2018)..
[270] E. Seo, H.M. Song, H.K. Kim, Gids: Gan based intrusion detection system for in-vehicle network, in: 2018 16th Annual Conference on Privacy, Security and Trust (PST), IEEE, 2018, pp. 1–6.
[271] P. Marek, V.I. Naik, V. Auvray, A. Goyal, Oodgan: Generative adversarial network for out-of-domain data generation, arXiv preprint arXiv:2104.02484 (2021)..
[272] Z. Zeng, H. Xu, K. He, Y. Yan, S. Liu, Z. Liu, W. Xu, Adversarial generative distance-based classifier for robust out-of-domain detection, in: ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 7658–7662.
[273] B. Xia, Y. Bai, J. Yin, Y. Li, J. Xu, Loggan: A log-level generative adversarial network for anomaly detection using permutation event modeling, Information Systems Frontiers 23 (2021) 285–298.
[274] B. Xia, J. Yin, J. Xu, Y. Li, Loggan: a sequence-based generative adversarial network for anomaly detection based on system logs, in: International Conference on Science of Cyber Security, Springer, 2019, pp. 61–76.
[275] M. Rigaki, S. Garcia, Bringing a gan to a knife-fight: Adapting malware communication to avoid detection, in: 2018 IEEE Security and Privacy Workshops (SPW), IEEE, 2018, pp. 70–75.
[276] M. Amin, B. Shah, A. Sharif, T. Ali, K. Kim, S. Anwar, Android malware detection through generative adversarial networks, Transactions on Emerging Telecommunications Technologies (2019) e3675.
[277] Y.-J. Zheng, X.-H. Zhou, W.-G. Sheng, Y. Xue, S.-Y. Chen, Generative adversarial network based telecom fraud detection at the receiving bank, Neural Networks 102 (2018) 78–86.
[278] A. Sethia, R. Patel, P. Raut, Data augmentation using generative models for credit card fraud detection, in: 2018 4th International Conference on Computing Communication and Automation (ICCCA), IEEE, 2018, pp. 1–6.
[279] D.M. Herr, B. Obert, M. Rosenkranz, Anomaly detection with variational quantum generative adversarial networks, Quantum Science and Technology (2021).
[280] J. Chen, Y. Shen, R. Ali, Credit card fraud detection using sparse autoencoder and generative adversarial network, in: 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, 2018, pp. 1054–1059.
[281] T. Leangarun, P. Tangamchit, S. Thajchayapong, Stock price manipulation detection using generative adversarial networks, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 2104–2111.
[282] T. Ali, S. Jan, A. Alkhodre, M. Nauman, M. Amin, M.S. Siddiqui, Deepmoney: counterfeit money detection using generative adversarial networks, PeerJ Computer Science 5 (2019) e216.
[283] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural computation 18 (2006) 1527–1554.
[284] G. Alain, Y. Bengio, L. Yao, J. Yosinski, E. Thibodeau-Laufer, S. Zhang, P. Vincent, Gsns: generative stochastic networks, Information and Inference: A Journal of the IMA 5 (2016) 210–249.
[285] Y.H. Tal Reiss, Mean-shifted contrastive loss for anomaly detection, 2021, arXiv:2106.03844.
[286] T. Reiss, N. Cohen, L. Bergman, Y. Hoshen, Panda–adapting pretrained features for anomaly detection, arXiv preprint arXiv:2010.05903 (2020)..
[287] F. Ye, H. Zheng, C. Huang, Y. Zhang, Deep unsupervised image anomaly detection: An information theoretic framework, 2020, arXiv:2012.04837.
[288] L. Bergman, N. Cohen, Y. Hoshen, Deep nearest neighbor anomaly detection, arXiv preprint arXiv:2002.10445 (2020)..
[289] K. Sohn, C.-L. Li, J. Yoon, M. Jin, T. Pfister, Learning and evaluating representations for deep one-class classification, International Conference on Learning Representations (2021)..
[290] Y. Chen, Y. Tian, G. Pang, G. Carneiro, Unsupervised anomaly detection with multi-scale interpolated gaussian descriptors, arXiv preprint arXiv:2101.10043 (2021)..
[291] J. Kim, K. Jeong, H. Choi, K. Seo, Gan-based anomaly detection in imbalance problems, in: European Conference on Computer Vision, Springer, 2020, pp. 128–145.
[292] V. Sehwag, M. Chiang, P. Mittal, Ssd: A unified framework for self-supervised outlier detection, in: International Conference on Learning Representations, 2021..
[293] Y. Liu, C. Zhuang, F. Lu, Unsupervised two-stage anomaly detection, 2021, arXiv:2103.11671.
[294] C. Huang, F. Ye, Y. Zhang, Y.F. Wang, Q. Tian, Esad: End-to-end deep semi-supervised anomaly detection, 2020, arXiv:2012.04905.
[295] J.T. Jewell, V.R. Khazaie, Y. Mohsenzadeh, Oled: One-class learned encoder-decoder network with adversarial context masking for novelty detection, arXiv preprint arXiv:2103.14953 (2021)..
[296] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, P. Gehler, Towards total recall in industrial anomaly detection, arXiv preprint arXiv:2106.08265 (2021)..
[297] D. Gudovskiy, S. Ishizaka, K. Kozuka, Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows, arXiv preprint arXiv:2107.12571 (2021)..
[298] V. Zavrtanik, M. Kristan, D. Skocaj, Draem-a discriminatively trained reconstruction embedding for surface anomaly detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8330–8339.
[299] T. Defard, A. Setkov, A. Loesch, R. Audigier, Padim: a patch distribution modeling framework for anomaly detection and localization, in: the 1st International Workshop on Industrial Machine Learning, ICPR 2020, 2020..
[300] H.M. Schlküter, J. Tan, B. Hou, B. Kainz, Self-supervised out-of-distribution detection and localization with natural synthetic anomalies (nsa), arXiv preprint arXiv:2109.15222 (2021)..
[301] J. Pirnay, K. Chai, Inpainting transformer for anomaly detection, arXiv preprint arXiv:2104.13897 (2021)..
[302] O. Rippel, P. Mertens, D. Merhof, Modeling the distribution of normal data in pre-trained deep features for anomaly detection, ICPR (2020).
[303] G. Wang, S. Han, E. Ding, D. Huang, Student-teacher feature pyramid matching for unsupervised anomaly detection, arXiv:2103.04257 (2021)..
[304] M. Rudolph, B. Wandt, B. Rosenhahn, Same same but differnet: Semi-supervised defect detection with normalizing flows, 2020, arXiv:2008.12577.
[305] J. Yi, S. Yoon, Patch svdd: Patch-level svdd for anomaly detection and segmentation, in: Proceedings of the Asian Conference on Computer Vision, 2020.
[306] V. Zavrtanik, M. Kristan, D. Skočaj, Reconstruction by inpainting for visual anomaly detection, Pattern Recognition 112 (2021) 107706.
[307] F.V. Massoli, F. Falchi, A. Kantarci, E. Akti, H.K. Ekenel, G. Amato, Mocca: Multi-layer one-class classification for anomaly detection, 2020, arXiv:2012.12111.
[308] N. Cohen, Y. Hoshen, Sub-image anomaly detection with deep pyramid correspondences, 2020, arXiv:2005.02357.
[309] D.-H.K. Jin-Hwa Kim, T.L. Saehoon Yi, Semi-orthogonal embedding for efficient unsupervised anomaly segmentation, arXiv:2105.14737 (2021)..
[310] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Computer Science (2014)..
[311] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[312] M. Tan, Q.V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, 2019..
[313] D. Jia, D. Wei, R. Socher, L.J. Li, L. Kai, F.F. Li, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
[314] Y. Xian, C.H. Lampert, B. Schiele, Z. Akata, Zero-shot learning–a comprehensive evaluation of the good, the bad and the ugly, IEEE transactions on pattern analysis and machine intelligence 41 (2018) 2251–2265.
[315] L. Feng, C. Zhao, Fault description based attribute transfer for zero-sample industrial fault diagnosis, IEEE Transactions on Industrial Informatics 17 (2020) 1852–1862.
[316] A.R. Rivera, A. Khan, I.E.I. Bekkouch, T.S. Sheikh, Anomaly detection based on zero-shot outlier synthesis and hierarchical feature distillation, IEEE Transactions on Neural Networks and Learning Systems (2020).

[317] M.H. Jarrahi, Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making, Business Horizons 61 (2018) 577–586.
[318] B. Kim, J. Gilmer, M. Wattenberg, F. Viégas, Tcav: Relative concept importance testing with linear concept activation vectors (2018)..
[319] C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye, A. Mordvintsev, The building blocks of interpretability, Distill 3 (2018) e10.
[320] E. Härkönen, A. Hertzmann, J. Lehtinen, S. Paris, Ganspace: Discovering interpretable gan controls, in: Proceedings of the International Conference on Neural Information Processing Systems, 2020.
[321] Y. Shen, C. Yang, X. Tang, B. Zhou, Interfacegan: Interpreting the disentangled face representation learned by gans, IEEE transactions on pattern analysis and machine intelligence (2020).
[322] Y. Shen, B. Zhou, Closed-form factorization of latent semantics in gans, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1532–1540.
[323] Y. Xu, Y. Shen, J. Zhu, C. Yang, B. Zhou, Generative hierarchical features from synthesizing images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4432–4442.
[324] S.K. Mustikovela, S. De Mello, A. Prakash, U. Iqbal, S. Liu, T. Nguyen-Phuoc, C. Rother, J. Kautz, Self-supervised object detection via generative image synthesis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 8609–8618.
[325] M.H. Bhuyan, D.K. Bhattacharyya, J.K. Kalita, Survey on incremental approaches for network anomaly detection, International Journal of Communication Networks and Information Security 3 (2011) 226–239.
[326] N. Tishby, F.C. Pereira, W. Bialek, The information bottleneck method, arXiv preprint physics/0004057 (2000)..
[327] Y. Du, J. Xu, H. Xiong, Q. Qiu, X. Zhen, C.G. Snoek, L. Shao, Learning to learn with variational information bottleneck for domain generalization, in: European Conference on Computer Vision, Springer, 2020, pp. 200–216.
[328] C. Jiang, Z. Zhang, Z. Chen, J. Zhu, J. Jiang, Third-person imitation learning via image difference and variational discriminator bottleneck (student abstract), in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 2020, pp. 13819–13820..
[329] A. v. d. Oord, O. Vinyals, K. Kavukcuoglu, Neural discrete representation learning, arXiv preprint arXiv:1711.00937 (2017)..
[330] P. Esser, R. Rombach, B. Ommer, Taming transformers for high-resolution image synthesis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12873–12883.
[331] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, I. Sutskever, Zero-shot text-to-image generation, arXiv preprint arXiv:2102.12092 (2021)..
[332] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, Transformers in vision: A survey, arXiv preprint arXiv:2101.01169 (2021)..
[333] N. Parmar, A. Vaswani, J. Uszkoreit, L. Kaiser, N. Shazeer, A. Ku, D. Tran, Image transformer, in: International Conference on Machine Learning, PMLR, 2018, pp. 4055–4064.

Xuan Xia received the Ph.D. degree in instrument science and technology from Shanghai Jiao Tong University, in 2017. He is currently a Researcher with the Shenzhen Institute of Artificial Intelligence and Robotics for Society. His research areas include deep learning, pattern recognition, image processing, time–frequency analysis, navigation and positioning, and signal processing.

Xizhou Pan received the master's degree in mechanical engineering from Shenzhen University, in 2020. He is currently an Assistant engineer with the Shenzhen Institute of Artificial Intelligence and Robotics for Society. His research interests include machine vision and deep learning.

Nan Li received his B.E. degree from Hunan University in 2011, and the Ph. D. degree in mechanical engineering from China Agricultural University in 2017. From 2015 to 2016, he was a visiting scholar with the Department of Agricultural and Biological Engineering, University of Illinois Urbana-Champaign. He was a postdoctoral researcher at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, and is currently working at Shenzhen Institute of Artificial Intelligence and Robotics for Society. His main research interests include machine vision and robotics.

Xing He received the B.E. from the College of Electronic Engineering, Guangxi Normal University in 2017, and the M.E. degrees from the College of Mechatronics and Control Engineering, Shen Zhen University in 2020. He is now working at Shenzhen Institute of Artificial Intelligence and Robotics for Society. His research interests include anomaly detection and deep learning.

Lin Ma received his B.S. degree from Xi 'an Jiaotong University in 1997, his honorary title of Microsoft Scholar in 2004, and his Ph.D. from Xi 'an Jiaotong University in 2009. He is currently working at Shenzhen Institute of Artificial Intelligence and Robotics for Society, focusing on the research of key technologies of computer vision and developmental cognitive learning.

Xiaoguang Zhang received the B.E. and M.E. degrees from the College of Information Engineering, Shen Zhen University, in 2017 and 2020. He is now working at Shenzhen Institute of Artificial Intelligence and Robotics for Society. His research interests include saliency detection and machine learning.

Ning Ding received the Ph.D. degree from the Department of Mechanic and Automation Engineering, The Chinese University of Hong Kong, Hong Kong SAR, in 2013. He is currently a Research Fellow and the Deputy Director of the Institute of Robotics and Intelligent Manufacturing, The Chinese University of Hong Kong, Shenzhen, Guangdong, China. His current research interests include bionic robot design, control, and computer vision.