Meta Pseudo Labels for Anomaly Detection via Partially Observed Anomalies

Sinong Zhao[1], Zhaoyang Yu[1], Xiaofei Wang[1], Trent G. Marbach[2], Gang Wang[1], and Xiaoguang Liu[1(✉)]

[1] College of Computer Science, NanKai-Orange D.T. Joint Lab, Nankai University, Tianjin, China
{zhaosn,yuzz,wangxf,wgzwp,liuxg}@nbjl.nankai.edu.cn
[2] Department of Mathematics, Toronto Metropolitan University, Toronto, Canada

Abstract. General anomaly detection based on weakly supervised or partially observed anomalies has been an important research. However, most such algorithms treat the unlabeled set as a substitute for normal samples and ignore the potential anomalies in it, which fails make full use of the abnormal supervision information. To address this issue, we propose a meta-pseudo-label based framework for anomaly detection (MPAD). The framework strives to obtain effective pseudo anomalies from the unlabeled samples to supplement the observed anomaly set. Specifically, a teacher network is improved based on the feedback of a student network on a validation set, thereby generating more conducive pseudo anomalies to assist the student network while incurring less confirmation bias. Extensive experiments show that the proposed MPAD algorithm outperforms current popular algorithms on five real datasets.

Keywords: Anomaly Detection · Semi-Supervised Learning · Meta Pseudo-Label

1 Introduction

Anomalies are generally defined as behaviors or events that are different from most normal situations which are rare but extremely harmful. Therefore accurate detection of anomalies are essential within many environments. Such environments may include fraud detection in finance [1], disease detection in clinical medicine [2], web intrusion detection [3] in network security, etc.

There have been many traditional anomaly detection algorithms based on unsupervised learning [4–6] or only normal class observed [7–9]. They usually assume that there are no observed anomalies during training and lose the chance to take advantage of the abnormal information. Consequently, a series of anomaly detection algorithms recently emerged that train models via a large number of unlabeled samples along with a few observed anomalies [10–12]. This setting is more in line with actual application scenarios, which can not only make up for the

X. Wang et al. (Eds.): DASFAA 2023, LNCS 13946, pp. 100–109, 2023.
https://doi.org/10.1007/978-3-031-30678-5_8

lack of supervision information in the unsupervised algorithms, but also reduce the burden of abnormal label collection in supervised schemes. Nevertheless, most of them directly regard unlabeled samples as a normal set, which may be unreasonable for some datasets containing a non-negligible amount of anomalies in unlabeled set. The core problem is finding out how to leverage the unlabeled samples to enhance the anomaly detection models.

Semi-supervised learning [13,14] is an appropriate choice to apply to anomaly detection due to its adequate mining of unlabeled samples. Some previous anomaly detection works [15,16] which are in semi-supervised frame simply applied unsupervised algorithms on unlabeled samples. In this paper, we employ a pseudo-label algorithm on the unlabeled set to find a set of *pseudo anomalies*.

Meta pseudo label (MPL) [17] takes the idea of meta-learning, which the teacher network continuously adjusts to reduce the confirmation bias using the feedback of the student network on the labeled samples. Inspired by MPL, we introduce a Meta-Pseudo-Label Anomaly Detection (MPAD) method in this paper. MPAD exploits the feedback of the student network on pseudo anomalies to influence the update of the teacher network. Meta pseudo anomalies (MPAs) then generated by the teacher network not only have less confirmation bias but also assist the student network to be more generalized on test set. In our implementation, we withhold a fixed validation set to judge the detection performance of the student network, and in turn the difference in performance is treated as a reward or punishment during the training of the teacher network.

The major contributions of this paper are summarized as follows:

- We propose an anomaly detection framework with partially observed anomalies which employs the pseudo-label algorithm to increase the content and quality of observed anomalies, thereby improving the accuracy of the anomaly detection model;
- The feedback of student model is used to correct the update direction of the teacher network, so that the teacher network can generate more beneficial pseudo anomalies;
- Extensive experimental results on datasets in five different fields show that the proposed MPAD framework exceeds five most currently popular algorithms in effectiveness.

2 Related Work

2.1 Anomaly Detection Methods

Traditional anomaly detection algorithms mainly follow the unsupervised setting. They cannot take advantage of existing anomaly information. Similar settings to this paper are semi-supervised or weakly-supervised based anomaly detection. One class of semi-supervised anomaly detection methods assumes that only normal samples are available when building a model. The classic algorithms are OCSVM [7] and deep support vector data description (SVDD) [8]. As they

only learn patterns of the normal category, any pattern that differs from the normal ones is considered as an anomaly. The advantage of this approach is that it can reduce the overfitting problem of abnormal learning. They generally assume that the data are similar within a class and they are mostly applicable to situations with a large number of positive samples. Another class of semi-supervised anomaly detection methods presumes that a small amount of labeled normal and anomalies are available in addition to unlabeled ones, e.g., DeepSAD [15] and the method in [16]. They are both based on SVDD. Generally speaking, these models outperform unsupervised algorithms due to the presence of supervised information. Some work [10–12,19,20] have the same detection settings as our MPAD and focus on a small number of observable anomalies and unlabeled samples. Yet, most of these works assume that unlabeled samples are normal, and our model extracts reliable pseudo anomalies from unlabeled samples to enhance the utilization of supervised information.

2.2 Semi-supervised Methods

At present, semi-supervised algorithms [13,14] are mainly based on consistency, pseudo-labels, and a class of hybrid algorithms. Consistency algorithms are mainly based on the assumption that different representations of the same sample can yield the same results on downstream tasks. Many of them rely on rich data augmentation. But pseudo-labels methods have no such problem. The meta pseudo label [17] method used the results of a student network on the labeled samples as the feedback to a teacher network, reducing the pseudo labels' confirmation bias. To the best of our knowledge, there are currently no anomaly detection algorithms based on partially observed anomalies that use pseudo-label algorithms. We propose a general framework of MPAD based on MPL, which can employ any network structure as the teacher and the student network, and is compatible with various types of data.

3 Methods

3.1 Preliminaries

We follow the setting that partially anomalies are observed in anomaly detection. Notationally, the dataset is represented by $\mathcal{D} = \{\mathcal{D}_{\mathrm{L}}, \mathcal{D}_{\mathrm{U}}\}$. $\mathcal{D}_{\mathrm{L}}$ notes the partially known anomalies set and $\mathcal{D}_{\mathrm{U}}$ is the unlabeled set in which normal samples are much more than anomalies. $\mathcal{D}_{\mathrm{L}} = \{(x_1, y_1), \cdots, (x_K, y_K)\}$, $\mathcal{D}_{\mathrm{U}} = \{x_{K+1}, \cdots, x_{K+N}\}$, where $x_i \in \mathcal{X}$, $\mathcal{X} = \mathbb{R}^d$, $y_i = 1$, $y_i \in \mathcal{Y}$, $\mathcal{Y} = \{0, 1\}$. Additionally, we follow the description of the models in pseudo-label algorithms. We define the teacher models that provide pseudo labels T, and their parameters θ_T. Student models that take pseudo labels, which in this paper are the anomaly detection models, are called S and the corresponding parameters are θ_S. We expect to train an anomaly detection model leveraging the dataset $\mathcal{D}$ and implement the model on the test set to examine its performance.

3.2 Meta Pseudo Anomaly Detection Scheme

We first introduce the basic pseudo-label algorithm, which obtains the distribution probability of the sample from a neural network, and gets the hard pseudo-label y^{PL} by a threshold λ:

$$y^{\mathrm{PL}} = \mathbb{1}\left[T(x_u;\theta_T) \geqslant \lambda\right], \tag{1}$$

in which $x_u \in \mathcal{D}_{\mathrm{U}}$ and $T(x_u;\theta_T)$ is the probability that x_u belongs to a particular class output by the teacher network. This formula is also used to generate pseudo anomalies when applied in anomaly detection.

This simplest pseudo-label method works well in anomaly detection, but the performance of student detection models are limited by the accuracy of the pseudo-labels produced by the teacher network. To improve this accuracy, we borrow the idea of Meta Pseudo Labels (MPL) to the pseudo anomalies generation which we called Meta Pseudo Anomalies (MPA).

We first utilize the teacher network to generate pseudo anomalies following Eq. (1). Here the teacher network refers to the self-training schedule, i.e., executing two steps in a loop: (1) Train a classifier using an already labeled dataset. Here we treat the unlabeled set as normal; (2) Use the trained classifier to label the unlabeled data, and add those with high prediction confidence to the labeled set. Based on these pseudo anomalies, the optimization objective θ_S^{MPA} of the student network is:

$$\theta_S^{\mathrm{MPA}} = \underset{\theta_S}{\operatorname{argmin}}\ \mathcal{L}_{\mathrm{CN}}\Big(S\big(\left[x_u, x_l, \mathrm{MPA}\right];\theta_S\big), y\Big), \tag{2}$$

where $x_l \in \mathcal{D}_{\mathrm{L}}$ are labeled anomalies and MPA is added to this set when training. $\mathcal{L}_{\mathrm{CN}}$ is the loss of student network presented in Eq. (10) and y is the true labels with the pseudo labels. So far this is a standard pseudo-label algorithm using self-training.

Seeing that the ultimate purpose of the student network is to improve the generalization effect on the test data. We expect the teacher network to generate pseudo anomalies that meet this goal. We manage to separate part of data called $\mathcal{D}_{\mathrm{V}}$ from $\mathcal{D}$ to do this. Since MPA is generated according to θ_T as in Eq. (1), the optimization result for student network can be seen as a function of θ_T which we write it as $\theta_S^{\mathrm{MPA}}(\theta_T)$. The overall goal is to minimize the loss of the student network on $\mathcal{D}_{\mathrm{V}}$:

$$\min_{\theta_S,\theta_T}\ \mathcal{L}_{\mathrm{CN}}\Big(S\big(\mathcal{D}_{\mathrm{V}};\theta_S^{\mathrm{MPA}}(\theta_T)\big), y_v\Big). \tag{3}$$

We expect that this objective will correct the update direction of the teacher network and further improve the performance of the detection network.

There are two variables in the target at the same time, so the parameters cannot be updated directly by calculating the derivative. Here we update the two parameters step-by-step depending on meta-learning. In order to achieve the approximate optimization, we let θ_T and θ_S update alternately. And only one step is updated each time along the gradient direction rather than directly

updating to the current optimal. This is because the current optimum is only a local optimum of the objective function according to the meta-learning theory. θ_S update one step to $\theta_S^{'}$ first:

$$\theta_S^{'} = \theta_S - \eta_S \nabla_{\theta_S} \mathcal{L}_{\mathrm{CN}}(\theta_T, \theta_S). \tag{4}$$

The contrastive above is applied with MPA generated by θ_T. θ_T is updated leveraging the updated student network $\theta_S^{'}$:

$$\theta_T^{'} = \theta_T - \eta_T \nabla_{\theta_T} \mathcal{L}_T(\theta_S^{'}). \tag{5}$$

We denote the objective function as H and split the derivative into a product of two derivatives:

$$\frac{\partial H}{\partial \theta_T} = \frac{\partial H}{\partial \theta_s^{'}} \cdot \frac{\partial \theta_s^{'}}{\partial \theta_T} = h \cdot \frac{\partial \mathcal{L}_{\mathrm{CE}}\big(\hat{y}_u, T(x_u; \theta_T)\big)}{\partial \theta_T}, \tag{6}$$

where $\hat{y}_u$ is the pseudo-labels and $h = \mathcal{L}_{\mathrm{CN}}(\theta_S) - \mathcal{L}_{\mathrm{CN}}(\theta_S^{'})$ following the Taylor's Formula. Both two contrastive losses are computed on the validation set. The second term in the last equation is the cross-entropy loss between the teacher network output and the pseudo-labels. In addition, we also trained the teacher network with the loss on the labeled samples. The total loss is as follows:

$$\mathcal{L}_T = \mathcal{L}_{\mathrm{CE}}\big(T(x_l; \theta_T), y_l\big) + \big(\mathcal{L}_{\mathrm{CN}}(\theta_S) - \mathcal{L}_{\mathrm{CN}}(\theta_S^{'})\big) \times \mathcal{L}_{\mathrm{CE}}\big(T(x_u; \theta_T), \hat{y}_u\big). \tag{7}$$

Here we assume that the unlabeled set is normal, and calculate the standard cross-entropy loss together with the labeled anomalies as the first term above.

3.3 Student Anomaly Learner

We chose DevNet [11] as the student model, which itself is a model based on a small number of observed anomalies. DevNet makes efficient use of observed anomalies. Its performance tends to increase on most datasets with the increase of observed anomalies. The main principles are: First, L abnormal scores of normal samples r_i are sampled from a standard Gaussian distribution, and the mean value is used as the reference score μ_r of the normal points:

$$\mu_r = \frac{1}{L} \sum_{i=1}^{L} r_i, r_i \sim \mathcal{N}(\mu = 0, \sigma = 1). \tag{8}$$

Then the z-score is applied to calculate the gap between the training data z_i and the reference score,

$$\mathrm{dev}(z_i) = \frac{z_i - \mu_r}{\sigma_r}. \tag{9}$$

Finally, the distance is increased between the abnormal points and the reference score while reducing the gap between the normal points and the reference score through the contrastive loss $\mathcal{L}_{\mathrm{CN}}$:

$$\mathcal{L}_{\mathrm{CN}} = (1 - y_i) \cdot |\mathrm{dev}(z_i)| + y_i \cdot \max\big(0, \delta - \mathrm{dev}(z_i)\big). \tag{10}$$

3.4 Total Flow of MPAD

We first initialize the teacher network and the student network with the observed anomalies and unlabeled samples, respectively. During training, batches are obtained from the initial dataset in a one-to-one ratio of normal and abnormal. The supervised loss of the teacher network is calculated within a batch. At the same time, a one-step update is made to the student network, and loss $(\mathcal{L}_{\mathrm{CN}}(\theta_S))$ on the validation set are recorded. The teacher network generates pseudo anomalies (MPA) according to the given probability threshold $\mathcal{P}^{\mathrm{MPA}}$. We add these MPAs to the observed anomalies set to secondly update the student network, and also record the loss $(\mathcal{L}_{\mathrm{CN}}(\theta_S^{'}))$ on the validation set. Finally, the teacher network is updated using the deviation of the loss on the student network and its own loss on labeled set.

4 Experiments

4.1 Experimental Settings

Datasets. We evaluate the proposed MPAD on five public datasets covering different fields.

Census is a dataset of US Census from 1994 and 1995 which includes 500 variables related to demographics and employment. Among them, very few people with an income more than 50,000 are regarded as anomalies for detection.

Campaign comes from a telemarketing campaign of a Portuguese bank. It contains 62 attributes such as customer information and economic activities. A small number of users who chose to subscribe to the banking product are identified as abnormal.

Thyroid is established to study whether patients had hypothyroidism. There are three categories which are normal, hyperfunctioning and dysfunctional. Here we merge the latter two categories as anomalies.

Arrhythmia is a dataset for studying arrhythmia and contains information about the patients' physical conditions and heart rates. Patients are classified into one normal class ECGs and 15 different types of arrhythmias. Here we combine the arrhythmia classes as anomalies.

Pima is a research dataset of diabetes in Pima Indian women, which comes from the UCI repository. Here we label those with diabetes as anomalies.

Baselines

- **DevNet** [11]: focuses on learning the anomaly scores directly rather than improving the representations. It designs a reference score of the normal samples according to the data distribution, and combines the contrastive loss to isolate the anomaly scores of normal samples and abnormal samples. It is an end-to-end anomaly detection algorithm based on partially observed anomalies and is also the student model of our MPAD.

- **DeepSAD** [15]: is a semi-supervised version of SVDD. It builds models with both unlabeled and labeled data. It places the normal samples close to the center of the hypersphere, while the abnormal samples are far from the surface of the hypersphere according to the label information, which improves the performance of SVDD.
- **SS-DGM** [21]: is a semi-supervised deep generative model. It combines a discriminative model of latent features with a generative semi-supervised model. This paper follows the setting of SS-DGM in [15] and applies it to anomaly detection.
- **OCSVM** [7]: is a classic single-class anomaly detection model which only use normal samples for training. It builds a hyperplane to segment samples, which maximize the separation between positive and negative samples.
- **iForest** [18]: is an efficient unsupervised model in anomaly detection. It achieves the isolation of anomalies by recursive segmentation of eigenvalues.

Implementation Details. We apply an MLP with a hidden layer as the teacher network in the implementation of MPAD. The architecture of the student model (DevNet) is the same as the teacher network, except that the teacher network outputs a two-dimensional vector, and the DevNet outputs a single-dimensional vector to calculate different losses. The number of neurons in the hidden layer is 64. In addition, the teacher network and the student network are optimized using the SGD and Adam optimizers, with learning rate of 0.03 and 0.001, respectively. The training and test sets of all algorithms are in a ratio of 8:2 with the random state of 42. DeepSAD, SS-DGM and the proposed MPAD are implemented with pytorch, while iForest and OCSVM are achieved with sklearn.

Metrics. AUC-ROC: is the area under the curve with the false positive rate as the abscissa and the true positive rate as the ordinate. It is a comprehensive evaluation criterion which represents the expected generalization performance of the model in different situations. Generally, if one curve can completely surround the other, it means that the former performs better than the latter, so the area under the curve is a good representation of the pros and cons of a model. In anomaly detection, it tends to show the ability to recognize normal classes due to extreme class imbalance.

AUC-PR: is the area under the curve drawn with the recall of the positive samples as the abscissa and the precision as the ordinate. It only pays attention to positive samples (anomalies). Similar to AUC-ROC, one curve is wrapped by another, indicating that the latter is more capable of achieving high recall and precision at the same time. In anomaly detection, we focus more on the detection ability of the anomaly category, so we care more about AUC-PR values than AUC-ROC values.

4.2 Effectiveness Results

The number of available anomalies in this comparative experiment is 30, and the noise of the training set is 0.02. Our MPAD and all baselines pick the best per-

Table 1. The performance w.r.t. AUC-ROC and AUC-PR among the proposed MPAD and the baselines on five tabular datasets with 30 labeled anomalies and 2% noise injection for training. The best performance for each dataset is boldfaced.

Datasets	AUC-ROC						AUC-PR					
	MPAD	DevNet	DeepSAD	SS-GDM	OCSVM	iForest	MPAD	DevNet	DeepSAD	SS-GDM	OCSVM	iForest
Census	**0.8906**	0.8284	0.7354	0.7683	0.5352	0.6165	**0.4118**	0.2895	0.0682	0.0420	0.0746	0.0744
Campaign	**0.8399**	0.8073	0.6337	0.7047	0.6942	0.7160	**0.4458**	0.3694	0.2334	0.2411	0.2530	0.2892
Thyroid	**0.9136**	0.8790	0.8957	0.7355	0.6153	0.7105	**0.6720**	0.3361	0.5921	0.4268	0.1225	0.2173
Arrhythmia	0.7557	0.7300	0.7751	**0.8107**	0.6844	0.7684	**0.5655**	0.4275	0.3912	0.4617	0.3774	0.5122
Pima	0.6845	0.7182	0.6698	**0.7598**	0.5487	0.6423	**0.6787**	0.6315	0.5788	0.5847	0.4378	0.4912

forming hyperparameters and demonstrate the optimal performance in Table 1. It can be seen that the proposed MPAD has achieved the best AUC-PR on all five datasets and the best AUC-ROC on three datasets. Among them, AUC-PR of MPAD exceeds the optimal result on each dataset by 12.2%, 7.6%, 8%, 5.3%, and 4.7%, respectively. This illustrates the advancement achieved by our algorithm. In general, the unsupervised algorithm iForest and the single-class model OCSVM do not perform as well as the first three baselines. Since DeepSAD and SS-GDM algorithms are also semi-supervised methods, they show the performance only second to MPAD on the *Thyroid* dataset. DevNet obtains the second position on the rest of the datasets. Among them, SS-GDM shows higher AUC-ROC on *Arrhythmia* and *Pima*, proving that it has better recognition of normal samples.

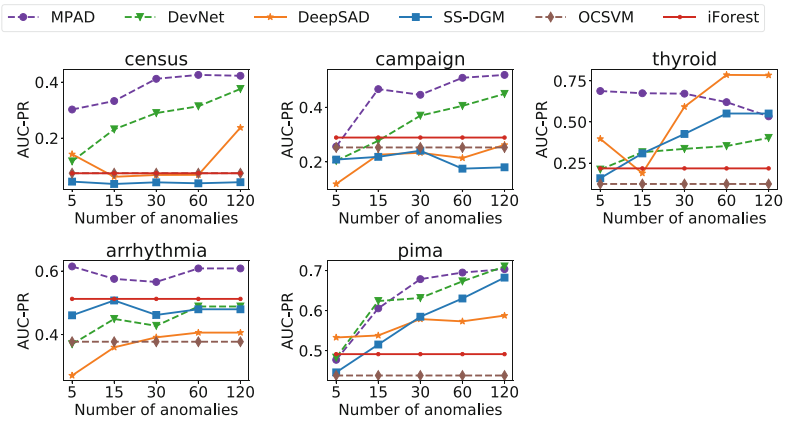

Fig. 1. AUC-PR w.r.t. No.labeled anomalies on five datasets.

4.3 Data Efficiency Study

This experiment aims to test how the performance of algorithms change as the observed anomalies increase. The noise ratio is fixed at 0.02 during this exper-

iment, and the observed anomalies are changed from 5 to 15, 30, 60, and 120 for modeling. It can be seen from Fig. 1 that our MPAD always maintains a high AUC-PR on *Census*, *Campaign*, and *Arrhythmia* datasets, and the results on *Pima* have a significant upward trend with the increase of observed anomalies. As for *Thyroid*, MPAD maintains a high level when there are fewer visible anomalies, and the effect decreases when the number of anomalies increases. The number of visible anomalies will act on the initialization effect of the teacher network. This result indicates that there are visible anomalies overlapping with the normal ones, making the teacher's performance drop further affecting the result of student. The two algorithms, OCSVM and iForest, are not influenced by the number of visible anomalies. They have certain advantages when there are few visible anomalies, yet they cannot make effective use of this information and lose their odds as the number of anomalies increases.

5 Conclusion

We introduce pseudo-label algorithms to partially-observed-anomalies anomaly detection. Thus, unlabeled data is used reasonably, and valuable pseudo anomalies can be extracted to assist the establishment of anomaly detection models. The most important part is that the proposed meta pseudo anomalies generation procedure makes the teacher network and the student network update alternately, and the teacher network is subject to both supervised information and the student network's feedback. Comprehensive experiments show that the pseudo anomalies generated in this way are better than the general pseudo labels, and our framework outperforms the other state-of-the-art anomaly detection methods on five public datasets.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (62272253, 62272252, 62141412) and Fundamental Research Funds for the Central Universities.

References

1. West, J., Bhattacharya, M.: Intelligent financial fraud detection: a comprehensive review. Comput. Secur. **57**, 47–66 (2016)
2. Fernando, T., Gammulle, H., Denman, S., Sridharan, S., Fookes, C.: Deep learning for medical anomaly detection-a survey. ACM Comput. Surv. (CSUR) **54**(7), 1–37 (2021)
3. Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J.: Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity **2**(1), 1–22 (2019). https://doi.org/10.1186/s42400-019-0038-7
4. Hoffmann, H.: Kernel PCA for novelty detection. Pattern Recogn. **40**(3), 863–874 (2007)
5. Chen, J., Sathe, S., Aggarwal, C., Turaga, D.: Outlier detection with autoencoder ensembles. In: Proceedings of the 2017 SIAM International Conference on Data Mining, pp. 90–98. SIAM (2017)

6. Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 665–674 (2017)
7. Li, K.L., Huang, H.K., Tian, S.F., Xu, W.: Improving one-class SVM for anomaly detection. In: Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693), vol. 5, pp. 3077–3081. IEEE (2003)
8. Ruff, L., et al.: Deep one-class classification. In: International Conference on Machine Learning, pp. 4393–4402. PMLR (2018)
9. Bergman, L., Hoshen, Y.: Classification-based anomaly detection for general data. arXiv preprint arXiv:2005.02359 (2020)
10. Zhang, Y.L., Li, L., Zhou, J., Li, X., Zhou, Z.H.: Anomaly detection with partially observed anomalies. In: Companion Proceedings of the The Web Conference 2018, pp. 639–646 (2018)
11. Pang, G., Shen, C., van den Hengel, A.: Deep anomaly detection with deviation networks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 353–362 (2019)
12. Pang, G., van den Hengel, A., Shen, C., Cao, L.: Toward deep supervised anomaly detection: reinforcement learning from partially labeled anomaly data. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 1298–1308 (2021)
13. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. arXiv preprint arXiv:2103.00550 (2021)
14. Chapelle, O., Scholkopf, B., Zien, A.: Semi-supervised learning (Chapelle, O. et al., EDS.; 2006)[book reviews]. IEEE Trans. Neural Netw. **20**(3), 542–542 (2009)
15. Ruff, L., et al.: Deep semi-supervised anomaly detection. arXiv preprint arXiv:1906.02694 (2019)
16. Görnitz, N., Kloft, M., Rieck, K., Brefeld, U.: Toward supervised anomaly detection. J. Artif. Intell. Res. **46**, 235–262 (2013)
17. Pham, H., Dai, Z., Xie, Q., Le, Q.V.: Meta pseudo labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11557–11568 (2021)
18. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data (TKDD) **6**(1), 1–39 (2012)
19. Pang, G., Cao, L., Chen, L., Liu, H.: Learning representations of ultrahigh-dimensional data for random distance-based outlier detection. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2041–2050 (2018)
20. Pang, G., Ding, C., Shen, C., van den Hengel, A.: Explainable deep few-shot anomaly detection with deviation networks. arXiv preprint arXiv:2108.00462 (2021)
21. Kingma, D.P., Mohamed, S., Jimenez Rezende, D., Welling, M.: Semi-supervised learning with deep generative models. In: Advances in Neural Information Processing Systems, vol. 27 (2014)