Self-Anomaly-Detection Model Training via Initialized Meta Model

Xindi MA*, Cunzhu SU*, Jianfeng MA*, Qi JIANG*, Ning XI*, Sheng Gao‡ and Kang XIE§¶
*School of Cyber Engineering, Xidian University, Shaanxi, China
‡School of Information, Central University of Finance and Economics, Beijing, China
§Key Lab of Information Network Security, Ministry of Public Security, Shanghai, China

2022 International Conference on Networking and Network Applications (NaNA) | 978-1-6654-6131-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/NANA56854.2022.00087

***Abstract*—Anomaly detection has become a key challenge affecting the training accuracy of machine learning. Because the training data is usually collected from Internet, many noised samples will be captured and these samples can decrease the model training accuracy. However, because the abnormal samples are difficult to predict when the samples are collected, and the training samples collected may contain many unknown exception categories, and the labels of normal samples may be incorrect, in this case, it is difficult to train an anomaly detection model based on supervised learning to accurately identify the anomaly samples. In this paper, we propose a new unsupervised anomaly detection method based on BiGAN, namely Rt-BiGAN, to identify the outliers in the training data. Firstly, we propose a Bigan network initialization method based on meta-learning algorithm with a small number of normal samples. Then, a self-supervised drop training is designed to improve the detection ability of the model. Finally, we evaluate our Rt-BiGAN over real-world datasets and the simulations results demonstrate that our mechanism is effective to detect the outliers in model training data.**

I. Introduction

With the development of big data, many computer vision jobs require enormous training datasets to achieve the deep learning model training, which is a form of data hunger [1]. To meet the data requirements, model trainer usually picks up the training data from the Internet to achieve classification and other visual research [2], [3], [4], [5]. In a typical Internet picture-driven application, we can look for photos on web search engines like Google and Bing or on photo-sharing sites, download a large number of images that correlate to a text query, and then model the target object/concept linked with the image collection. Search engine or site crawler images, on the other hand, usually captures some noised samples, which can impact the learning model. As a result, the outliers, or irrelevant images, in the collected training data need to be removed.

Anomaly detection is one of the technique to detect the abnormal, unreasonable or even wrong elements from data or events. Because the abnormal samples are difficult to predict when the samples are collected and the training samples collected may contain many unknown exception categories, the traditional anomaly detection methods based on supervised learning are difficult to accurately identify the abnormal samples. Thus, most of the existed works regarding the detection as a single classification task, have limited application in the practical anomaly detection. In other aspects, unsupervised learning methods based on reconstruction, such as [10], [11], [12] , and [13] , have received a lot of attention to achieve outlier detection recently.

As one of the techniques to generate samples, Generative Adversarial Networks (GAN) has been widely used to achieve the anomaly detection, such as AnoGAN[17], f-AnoGAN[18], EBGAN[19], GANomaly[20],and MAD-GAN[21]. The aim of these methods are to train GAN with normal sample data so that the generator can learn the distribution of normal data completely. Therefore, the trained GAN has the ability to reconstruct normal data samples. And when the input data is an abnormal sample that has never been seen before, the reconstructed sample and the original input sample will show significant differences. Some form of residuals between the actual instance and the generated instance are defined as the exception score. When the exception score exceeds a certain range, this sample is regarded as an outlier sample. But the anomaly detection methods based on GAN are meaningful only if there is no contamination in the training dataset, which means there are only normal samples in the training set. When the training set is contaminated by abnormal samples, the GAN model can not only capture the distribution of normal data, but also capture the distribution of abnormal data.

In order to solve these problems, we propose a new unsupervised anomaly detection method based on BiGAN, called Rt-BiGAN. The intuition behind this is to initialize the BiGAN model with a small number of normal samples (usually 5-10 images) chosen from the training set. After the initialization of BiGAN network on the self-supervised iterative training, we further improve the detection ability during the model training. In summary, the main contribution of this paper is as follows:

1) An initialization method of BiGAN network is proposed, which combines the meta-learning to train the BiGAN mode over a small number of samples. Then training a model with preliminary detection capability on this meta-model requires only a small amount of normal data. Therefore, even if the training set is contaminated by abnormal data, we can get a small amount of normal images from the contaminated data set through manual or other training. That is, an anomaly detection model can be trained on a contaminated dataset.

¶Kang XIE is corresponding author. Email: xiekang@stars.org.cn

978-1-6654-6131-3/22/$31.00 ©2022 IEEE
DOI 10.1109/NaNA56854.2022.00087

2) BiGAN detection model trained by a small amount of normal data has limited detection capability. Thus, we propose to carry out self-supervised drop training on the contaminated data set to improve the detection capability of the model.
3) Finally, we evaluate Rt-BiGAN over real-world dataset and the simulation results demonstrate that our Rt-BiGAN can effectively detect the outlier samples in training dataset.

II. Related Work

There are many researches on unsupervised anomaly detection, and the existing methods can be divided into three categories: 1)cluster-based detection, 2)classification-based detection, and 3) reconstruction-based detection.

A. Cluster-based Detection

The anomaly detection method based on clustering is a common unsupervised method[5], [6]. This method is based on the local assumption that the normal data belong to the majority, and the isolated point deviating from most data is regarded as the most likely outlier. K-means and DB-scan are the classic algorithms in this field. However, the detection accuracy of this class is relied on the clustering result, and some clustering algorithms were very sensitive to outliers. Thus, when there is contaminated data in the given data, clustering and subsequent anomaly detection performance would be misleading.

B. Classification-based Detection

The goal of these algorithms is to learn specific feature representations for anomaly detection using one-class classification. The problem solved by this method is to learn the description of a set of data instances to evaluate whether the new instance conforms to the training data. Support Vector Machine(SVM) influenced most one-class classification models, such as one-class SVM[7] and Support Vector Data Description (SVDD)[8]. Although these methods are simple to comprehend and use, they frequently perform poorly as the dimensions expand due to dimensional catastrophes.

C. Reconstruction-based Detection

By assuming the small proportion of the outlier samples in the original data, the outlier data can not be effectively reconstructed from the low-dimensional representation. The abnormal fraction is usually defined as the reconstruction error, and the larger the reconstruction error is, the greater the probability that the detected sample is abnormal. The classical method of this kind is PCA[9]. For more complex data, a deep learning model based on PCA has been proposed[10]. The model based on deep auto-encoder[14], [15], [16] lacks an effective regularization method for anomaly detection, while the model based on GAN introduces antagonistic regularization to alleviate over-fitting. However, when the training set is contaminated by abnormal samples, the traditional GAN-based[17], [18], [19], [20], [21] model can not only capture the distribution of normal data, but also capture the distribution of abnormal data.

III. Preliminaries

A. BiGAN Network

BiGAN network consists of three parts: Generator $\mathtt{G}$, encoder $\mathtt{E}$, and discriminator $\mathtt{D}$. Generator $\mathtt{G}$ maps the noise to a fake sample, while encoder $\mathtt{E}$ maps the real sample to a hidden variable space. Encoder $\mathtt{E}$ and generator $\mathtt{G}$ must learn to invert each other to deceive discriminator $\mathtt{D}$. BiGAN's objective function is as follows,

$$V(\mathtt{D},\mathtt{E},\mathtt{G}) = E_{x\sim p_x}[E_{z\sim p_E(.|x)}[logD(x,z)]]+ \\ E_{z\sim p_z}[E_{x\sim p_G(.|z)}[1-logD(x,z)]] \quad (1)$$

Here,$p_x(x)$ is the distribution over the data, $p_z(z)$ is the distribution over the latent representation,and $p_E(x|z)$ and $p_G(x|z)$ are the distributions induced by the encoder and generator respectively. Compared to many gan-based anomaly detection methods such as AnoGAN, BIGAN can learn not only the distribution of input data, but also the encoder to map input sample X to potential representation Z, which allows us to avoid the computatively expensive step of restoring a potential representation during testing.

B. Meta Learning

Meta-learning is a process of learning how to learn. The meta-learning algorithm is a composition of several learning tasks. Each task is a learning problem, and outputs a fast learner, which can learn and generalize from a small number of samples. MAML is by far the most widely used meta-learning algorithm. Like MAML, Reptile is a parameter initialization method for learning the neural network so that it can be adjusted with a small amount of new task data. However, compared to MAML, the Reptile algorithm is simpler and faster to train for new tasks.

IV. Problem Formulation

As shown in Fig. 1, central server (CS) is a globally trusted server that trains a target model, such as a cat and dog classification model, based on data uploaded by each client. Each client has its own unique data in their local location. These clients upload their own data to a central server, and the data is not shared between the clients. Because the central server has no restriction on the data uploaded by the client,this results in training the model using task-independent data(anomaly data), such as the face and car images shown in Fig. 1, resulting in a loss of model accuracy.

We are given a large training dataset D of m size, $D = (x_1, ..., x_m)$ containing both normal and abnormal images , and a small test data set $D = (x_1, y_1), ..., (x_n, y_n)$, where $y_i \in [0, 1]$ represents the image label and $m >> n$. Given a dataset, our goal is to first model the normal data in D to learn its manifold, and then detect the abnormal samples in D as outliers in the inference phase. Model F learns the normal data distribution and minimizes the outlier score $A(x)$. For a given test image x, $A(x)$ with a high outlier indicates that there may be an anomaly in the image. This is evaluated by setting a threshold ϕ for the score, where $A(x) > \phi$ represents the anomaly.

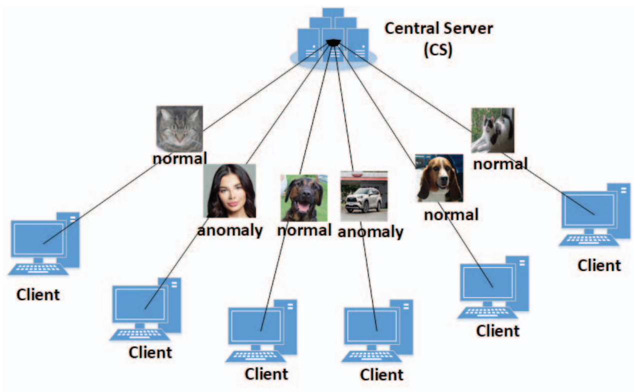


Fig. 1: System Model

V. Details of Rt-BiGAN

Efficient-GAN (Houssam Zenati e al. , 2019)[19] is an anomaly detection model based on BiGAN, which can be used for fast detection by learning an encoder E that maps an input sample X to a potential representation of Z. This method is the same as the GAN-based method. By learning the distribution of the training samples, the model can give a higher outlier score to the outliers deviating from the normal distribution. In the detection phase, the outlier fraction is calculated by the following formula,

$$A(x) = \alpha L_G(x) + (1-\alpha) L_D(x) \tag{2}$$

Here, $L_G(x) = ||x - G(E(x))||_1$ and $L_D(x)$ is the feature-matching loss.Based on $A(x)$, we can compare it with the preset threshold value to judge whether the sample is abnormal or not. Unfortunately, the models with millions of parameters tend to learn too much and over-adaptive to rare training samples.Therefore, we suggest that instead of directly using training data (both normal and abnormal) to train the model, a small amount of normal data should be culled from the training data to initialize the detection model. Then, using the initialized model, self-monitoring training is carried out on the training data to improve the detection ability.

As shown in Fig. 2, our proposed approach consists of two main steps. The first step is to combine the meta-learning with BiGAN and train the BiGAN network on multiple tasks to get a meta-model with low loss on each task. After the meta-training, we take a few pieces of normal data on the meta-model to update several adaptive steps and obtain a BiGAN model with preliminary anomaly detection ability. The second step is to input all the training data into the initialization detection model and get the outliers of all the data. Then the model is trained with the data of low outlier score, and the self-training iterative learning algorithm is used to improve the anomaly detector.

A. Initializing of Anomaly Detection

The role for the initializing of anomaly detection is to obtain a BiGAN model with preliminary detection ability that train data can be identified as belonging to A(anomaly) and N(normal) with high probability. To achieve this, we first use meta-learning algorithm to learn the initialized parameters of BiGAN Network, and then use a small amount of normal data to fine-tune the model, which can obtain a BiGAN model with anomaly detection ability.

1) Rt-BiGAN: BiGAN can be trained in unsupervised manner. But this kind of method needs a lot of data support. The Reptile algorithm is suitable for small-sample learning and can learn the transferable internal representations among various meta-tasks. Therefore, Reptile learning is introduced into BiGAN and a data generation algorithm Rt-BiGAN is proposed. Rt-BiGAN trains the optimal initialization parameters to converge rapidly on the new task (small sample data generation task), and gets the specific BiGAN. However, the BiGAN network trained in this article is not to achieve a small sample of data generation expansion, but for anomaly detection.

2) Base-Learners: Base-learners W is inherited from meta-learners, and its model is composed of encoder E, generator G and discriminator D. Encoder E is a mapping of data to noise, generator G is a mapping of noise to data, and discriminator D is a mapping of data and noise tuples to true and false classes. We denote encoder parameters as W_e, generator parameters as W_g, discriminator parameters as W_d. The base-learners are trained on a set of meta-tasks, whose objective function is in formula 1.

3) Meta-Learners: The meta-learner model Φ is also composed of encoder E, generator G and discriminator D. We also denote encoder parameters as Φ_e, generator parameters as Φ_g, discriminator parameters as Φ_d. Base learners can only learn the data characteristics of the current task, not suitable for other tasks. The purpose of the meta-learner is to balance the learning of the basic learners effects, and find the optimal initialization model for all tasks.

4) Rt-BiGAN Training Strategy: The whole Rt-BiGAN training process can be shown in Algorithm 1. The training process includes two parts: internal circulation and external circulation. In the outer loop, each basic-learner initializes the parameters by copying the meta-learner's parameters. In the inner-loop part, the basic-learner updates its parameters on task t by k-step counterwork training, and the new parameters are recorded as W_t. At the end of the inner loop, update the meta-learner parameter by setting the gradient to $\Phi - W_t$.

After training, we use a similar process to generate novel images from the sampled class described in Algorithm 2.

After a few steps of gradient updating on the normal data set, we can get a model that can detect outliers. However, the ability of this detection model is limited and the generalization ability is low. Thus, we propose to continue the self-supervised training in the training set, which contains both normal data and abnormal data.

B. Iterative Learning via Self-Training

Self-training, also known as self-learning, is a simple and effective semi-supervised learning framework, which has been widely used in many practical applications. In the process of

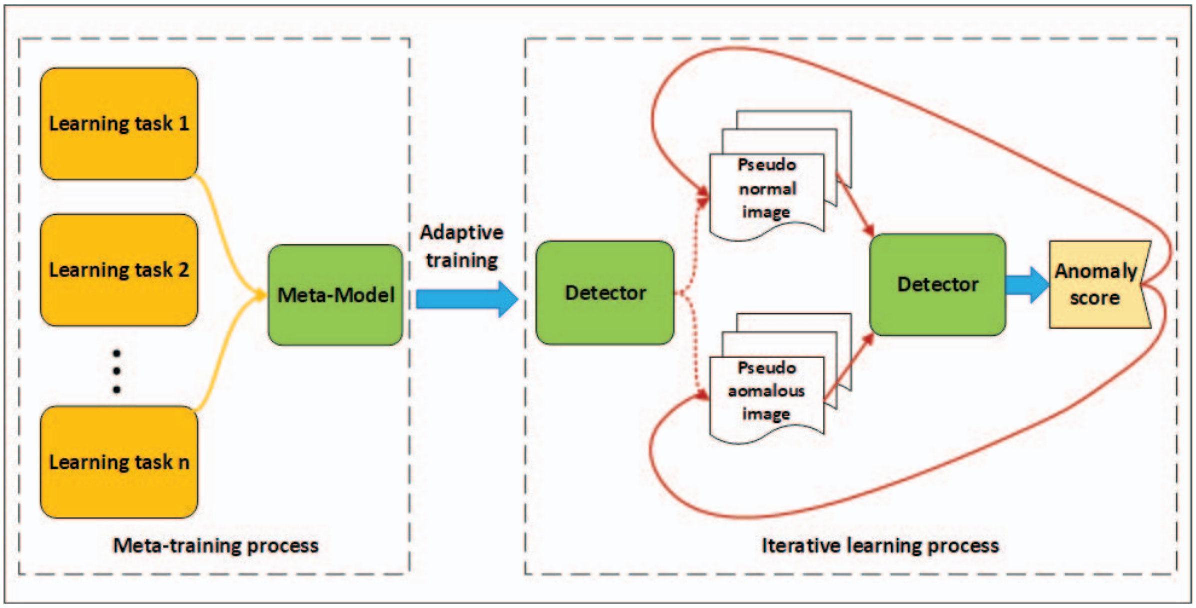


Fig. 2: The proposed framework. Given a set of unlabeled images, we first perform BiGAN network generation training on multiple tasks, and then do adaptive training on meta-model to get an anomaly detection model.The detection model is then used to recompute the anomaly scores of all images. The membership of A and N is updated accordingly, and the process repeated.

Algorithm 1 Rt-BiGAN Training

Initialize Φ_e,the encoder parameters for meta-learners
Initialize Φ_d,the discriminator parameters for meta-learners
Initialize Φ_g,the generator parameters for meta-learners
for $iteration$ 1,2,3... **do**
  Make a copy of Φ_e resulting in W_e
  Make a copy of Φ_d resulting in W_d
  Make a copy of Φ_g resulting in W_g
  Sample n images from X_t resulting x_t
  for $K > 0$ **do**
    Generate latent vector z
    Generate fake images y with z and W_g
    Generate fake latent vector z' with x_t and W_e
    Update W_d to increase $discriminator(x_t, z')$ and decrease $discriminator(y, z)$
    Update W_g and W_e to decrease $discriminator(x_t, z')$ and increase $discriminator(y, z)$
    $K \Leftarrow K - 1$
  end for
  Set Φ_e gradient to be $\Phi_e - W_e$
  Perform step of Adam update on Φ_e
  Set Φ_d gradient to be $\Phi_d - W_d$
  Perform step of Adam update on Φ_d
  Set Φ_g gradient to be $\Phi_g - W_g$
  Perform step of Adam update on Φ_g
end for

Algorithm 2 Rt-BiGAN Generation

Using W_e,a copy of the meta-trained Φ_e
Using W_d,a copy of the meta-trained Φ_d
Using W_g,a copy of the meta-trained Φ_g
Sample test task t
Sample n images as x_t from X_t
for $K > 0$ **do**
  Generate latent vector z
  Generate fake images y with z and W_g
  Generate fake latent vector z' with x_t and W_e
  Update W_d to increase $discriminator(x_t, z')$ and decrease $discriminator(y, z)$
  Update W_g and W_e to decrease $discriminator(x_t, z')$ and increase $discriminator(y, z)$
  $K \Leftarrow K - 1$
end for
Generate latent vector z with W_e and test image x
Generate fake image y with W_g and latent vector z

self-training, a small amount of labeled data is used to train the classification model. The model is then used to predict the classes of unlabeled data. These pseudo labels are combined with labeled data to train a new classification model. The prediction is then repeated and the new observations are added to the pseudo-label pool. Iterate over these steps until the maximum number of iterations specified is reached.

We obtain the initialized parameters for the detection model in the previous step. Then, we input all the training data including normal data and abnormal data into the initialized detection model to get their reconstructed images. According to Formula (2), the abnormal scores of the unlabeled samples were calculated and assigned with pseudo-labels, that is, the abnormal scores below the threshold were labeled as '0' (normal samples), and the abnormal scores above the threshold were labeled as '1'(abnormal samples). The new normal sample set is used to update the detector to improve the ability of the detection model. The entire training process is iterated until a predefined stopping criterion is met.In the iterative training process, we differ from the traditional confidence-based method to select reliable samples. Instead of adding more label data, we replace the last round of labels with the newly acquired labels and retrained the abnormal detectors. The main reason for this is that in the absence of monitoring information, the combination of the previous pseudo-label with the newly acquired pseudo-label can lead to worse labels. Based on experiments, we find that this simple strategy is very effective.

VI. Experiments

A. Datasets

In the evaluation, we adopt MNIST and Fashion-MNIST datasets to evaluate our Rt-BiGAN. MNIST dataset contains 60,000 samples as training data and 10,000 samples as test data. Fashion-MNIST also contains 60,000 images in 10 categories. We set the MNIST dataset as normal samples and Fashion-MNIST as abnormal samples. Thus, while the model training, we need to identify the Fashion-MNIST samples from the model training dataset.

B. Performance Evaluation Metrics

Area Under Curve (AUC) is defined as the Area Under the ROC Curve enclosed by the coordinate axis. The ROC Curve is called receiver operating characteristic Curve, which is based on a series of different binary classification of values or decisions. AUC is a kind of performance index to measure the performance of two-classification learners. Thus we use AUC as a performance evaluation index in our experiment. By definition, AUC can be obtained by summing the areas under the ROC curve.

C. Image Generation Experiments

Fig. 3 is the result of Rt-BiGAN generating an invisible test class on the Mnist dataset (100,000 iterations on the training task).The first line of each number represents the training data. The next three lines are the model-generated images (after fine-tuning 20 gradient steps on these data points)

D. Effective in Mnist Dataset

As shown in Fig. 4 .(a) , we first train the BiGAN network on a training dataset containing only normal data and get the ROC curve on the test set to test its performance. It can be seen that the detection performance of the model is high without abnormal data, and its AUC value is up to 0.98. We then train the anomaly detection model on the dataset

Fig. 3: Generated Samples

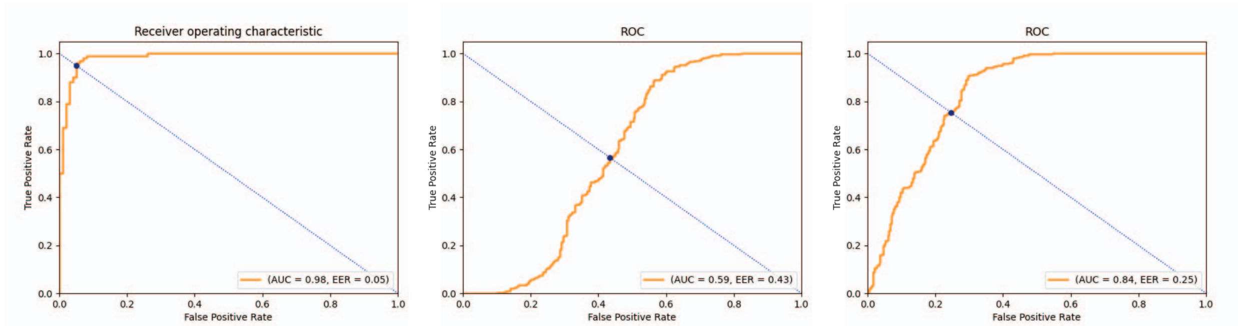
containing the anomaly data, and also get the ROC curve on the dataset. As shown in Fig. 4 .(b), since the training dataset contains abnormal data and the BiGAN network has strong learning ability, the detection performance of the anomaly detection model decreases significantly by learning not only the distribution of normal data but also the distribution of abnormal data. The ROC curve is significantly lower than the model trained on the normal data set, the AUC is only 0.59, which is a 0.33 drop between 25 and 35, and it is also significantly lower than the model trained using only normal data. Finally, we evaluate the anomaly detection performance of the model by semi-supervised iterative training on the initial model. The results are shown in Fig. 4 .(c). It can be seen that the ROC curve of our method is slightly less than that of the model trained with only normal data and its AUC is 0.84. Compared with the model trained over abnormal training data, the AUC value is increased from 0.59 to 0.84, and the AUC value is increased by 0.25. The performance is improved obviously.At the same time, the AUC value of the initial anomaly detection model is about 0.6, which can be increased by 10-15% through iterative training. This shows that the self-training method can improve the performance of the detection model iteratively.

TABLE I: AUC performance of different methods on normal dataset and polluted dataset.

Method	Normal data	Polluted data
AnoGAN	0.99	0.46
EGBAD	0.98	0.52
Rt-BiGAN	**0.98**	**0.84**

E. Performance comparison

We trained Anogan, EGBAD, and our model Rt-BiGan on normal datasets 1 containing only mnist data and contaminated datasets 2 containing both mnist and fashion data, finally, it is detected on the test set. The results are shown in Table. I , and it is clear that our approach works better on thie contamination data set than other approaches.



(a) Normal training set (b) Polluted training set (c) Polluted training set

Fig. 4: The ROC curve of the algorithm

VII. Conclusion

In this paper, we study an anomaly detector in federated learning. We proposed a tiruly unsupervised anomaly detection data management method to maintain the results even if the training data is polluted. The anomaly detection model was initialized by training a BiGAN meta-model on several generating tasks, and then we feed some normal data to perform several steps of gradient tracking on the meta-model. Finally, a semi-supervised self-training is carried out based on the initial anomaly detection model to further improve the detection capability. The simulation results demonstrated that our mechanism can identify the outliers effectively.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61902290, 62072352, U21A20464, 61872283, 62072487),Key Research and Development Program of Shaanxi(Grant Nos. 2020ZDLGY09-06, 2019ZDLGY12-04), Key Lab of Information Network Security, Ministry of Public Security (Grant No. C19604), Scientific Research Program Funded by Shaanxi Provincial Education Department (Grant No. 20JY016), Fundamental Research Funds for the Central Universities (Grant Nos. XJS211515).

References

[1] Yan Xia, Xudong Cao, Fang Wen, Gang Hua and Jian Sun. Learning Discriminative Reconstructions for Unsupervised Outlier Removal (2015).

[2] R. Fergus, L. Fei-Fei, P . Perona, and A. Zisserman. Learning object categories from internet image searches. Proceedings of the IEEE, 98(8):1453–1466, 2010.

[3] R. Fergus, L. Fei-Fei, P . Perona, and A. Zisserman. Learning object categories from internet image searches. Proceedings of the IEEE, 98(8):1453–1466, 2010.

[4] W. Liu, Y .-G. Jiang, J. Luo, and S.-F. Chang. Noise resistant graph ranking for improved web image search. In Proc. CVPR, 2011.

[5] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. 2018. Deep clustering for unsupervised learning of visualfeatures. In ECCV. 132–149.

[6] Nat Dilokthanakul, Pedro AM Mediano, Marta Garnelo, Matthew CH Lee, Hugh Salimbeni, Kai Arulkumaran, and Murray Shanahan.2017. Deep unsupervised clustering with gaussian mixture variational autoencoders. In ICLRp. 224, 2017.

[7] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. 2001. Estimating the support of a high-dimensional distribution. Neural Computation 13, 7 (2001), 1443–1471

[8] David MJ Tax and Robert PW Duin. 2004. Support vector data description. Machine Learning 54, 1 (2004), 45–66.

[9] D. H. Hoang and H. D. Nguyen, "A pca-based method for iot network traffic anomaly detection," in 2018 20th International conference on advanced communication technology (ICACT). IEEE, 2018, pp. 381–386.

[10] N. Merrill and C. C. Olson, "Unsupervised ensemble-kernel prin- cipal component analysis for hyperspectral anomaly detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 112–113.

[11] H. Alshammari, O. Ghorbel, M. Aseeri, and M. Abid, "Non- negative matrix factorization (nmf) for outlier detection in wireless sensor networks," 506–511.

[12] H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y . Liu, Y . Zhao, D. Pei, Y . Feng et al., "Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications," in Proceedings of the 2018 World Wide Web Conference, 2018, pp. 187–196.

[13] K. Hundman, V . Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in Proceedings of the 24th

[14] Carl Doersch. 2016. Tutorial on variational autoencoders. arXiv preprint:1606.05908 (2016).

[15] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research 11, Dec (2010), 3371–3408.

[16] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. 2011. Contractive auto-encoders: explicit invariance during feature extraction. In ICML. 833–840.

[17] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In IPMI. Springer, Cham, 146–157.

[18] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. 2019. f-AnoGAN: Fast unsuper-vised anomaly detection with generative adversarial networks. Medical Image Analysis 54 (2019), 30–44.

[19] Houssam Zenati, Chuan Sheng Foo, Bruno Lecouat, Gaurav Manek, and Vijay Ramaseshan Chandrasekhar. 2018. Efficient gan-based anomaly detection. arXiv preprint:1802.06222 (2018).

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

[21] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "Mad-gan:Multivariate anomaly detection for time series data with genera-tive adversarial networks," in International Conference on Artificial Neural Networks. Springer, 2019, pp. 703–716.