ICML
International Conference
On Machine Learning
2023 Fortieth International Conference on Machine Learning

Fascinating Supervisory Signals and Where to Find Them: Deep Anomaly Detection with Scale Learning

Hongzuo Xu[1], Yijie Wang[1], Juhui Wei[2], Songlei Jian[1], Yizhou Li[1], Ning Liu[1]

[1]National Key Laboratory of Parallel and Distributed Computing, College of Computer
[2]College of Science
National University of Defense Technology, China

Honolulu, Hawai'i, USA

July 2023

Motivation: Background

- Deep anomaly detection methods yield drastic performance improvement over traditional methods

Ensemble of Autoencoders

AUC-ROC of Thyroid:
0.90 (RandNet) vs. 0.63, 0.44, and 0.73

Data set	*RandNet*	Hawkins [10]	LOF [6]	HiCS [9]	Spectral [18]
Ca[illegible]	[illegible]	[illegible]	50.63	92.37	78.90
Ecoli	85.42	82.87	39.35	53.89	91.81
Ly[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	78.16
Optdigits	87.11	**87.63**	67.11	43.63	2.66
Pendigits	**93.44**	89.81	54.37	60.61	87.88
Seismic	**71.28**	68.25	55.59	59.90	66.71
Thyroid	**90.42**	87.47	63.04	43.76	72.94
Waveform	**70.05**	61.57	55.48	59.24	62.88
Yeast	**82.95**	82.12	54.30	54.45	77.70

Chen, J., Sathe, S., Aggarwal, C., and Turaga, D. **Outlier detection with autoencoder ensembles**. In SIAM International Conference on Data Mining, pp. 90–98. SIAM, 2017.

Deep SVDD

Consistent Superiority

Normal Class	OC-SVM/ SVDD	KDE	IF	DCAE	AnoGAN	Soft-Bound. Deep SVDD	One-Class Deep SVDD
0	**98.6**±0.0	97.1±0.0	98.0±0.3	97.6±0.7	96.6±1.3	97.8±0.7	98.0±0.7
1	99.5±0.0	98.9±0.0	97.3±0.4	98.3±0.6	99.2±0.6	99.6±0.1	**99.7**±0.1
2	82.5±0.1	79.0±0.0	88.6±0.5	85.4±2.4	85.0±2.9	89.5±1.2	**91.7**±0.8
3	88.1±0.0	86.2±0.0	89.9±0.4	86.7±0.9	88.7±2.1	90.3±2.1	**91.9**±1.5
4	[illegible]	[illegible]	[illegible]	[illegible]	89.4±1.3	93.8±1.5	**94.9**±0.8
5	[illegible]	[illegible]	[illegible]	[illegible]	88.3±2.9	85.8±2.5	**88.5**±0.9
6	96.5±0.0	87.6±0.0	95.6±0.3	94.6±0.5	94.7±2.7	98.0±0.4	**98.3**±0.5
7	93.7±0.0	91.4±0.0	92.0±0.4	92.3±1.0	93.5±1.8	92.7±1.4	**94.6**±0.9
8	88.9±0.0	79.2±0.0	89.9±0.4	86.5±1.6	84.9±2.1	92.9±1.4	**93.9**±1.6
9	93.1±0.0	88.2±0.0	93.5±0.3	90.4±1.8	92.4±1.1	94.9±0.6	**96.5**±0.3
AIRPLANE	61.6±0.9	61.2±0.0	60.1±0.7	59.1±5.1	**67.1**±2.5	61.7±4.2	61.7±4.1
AUTOMOBILE	63.8±0.6	64.0±0.0	50.8±0.6	57.4±2.9	54.7±3.4	64.8±1.4	**65.9**±2.1
BIRD	50.0±0.5	50.1±0.0	49.2±0.4	48.9±2.4	**52.9**±3.0	49.5±1.4	50.8±0.8
CAT	55.9±1.3	56.4±0.0	55.1±0.4	58.4±1.2	54.5±1.9	56.0±1.1	**59.1**±1.4
DEER	66.0±0.7	**66.2**±0.0	49.8±0.4	54.0±1.3	65.1±3.2	59.1±1.1	60.9±1.1
DOG	62.4±0.8	62.4±0.0	58.5±0.4	62.2±1.8	60.3±2.6	62.1±2.4	**65.7**±2.5
FROG	74.7±0.3	**74.9**±0.0	42.9±0.6	51.2±5.2	58.5±1.4	67.8±2.4	67.7±2.6
HORSE	62.6±0.6	62.6±0.0	55.1±0.7	58.6±2.9	62.5±0.8	65.2±1.0	**67.3**±0.9
SHIP	74.9±0.4	75.1±0.0	74.2±0.6	**76.8**±1.4	75.8±4.1	75.6±1.7	75.9±1.2
TRUCK	75.9±0.3	**76.0**±0.0	58.9±0.7	67.3±3.0	66.5±2.8	71.0±1.1	73.1±1.2

Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., Muller, E., and Kloft, M. **Deep one-class classification**. In Proceedings of the 35th International Conference on Machine Learning, volume 80, pp. 4393–4402, 2018.

Deep Isolation Forest

Deep iForest achieves 50%+ AUC-PR improvement over iForest and its shallow extensions

AUC-PR				
DIF (ours)	EIF	PID	LeSiNN	IF
0.404$\pm$0.051	0.198$\pm$0.022	0.075$\pm$0.007	0.183$\pm$0.028	0.063$\pm$0.006
0.453$\pm$0.051	0.218$\pm$0.028	0.066$\pm$0.005	0.205$\pm$0.031	0.046$\pm$0.004
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]$\pm$0.005
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]$\pm$0.003
[illegible]	[illegible]	[illegible]	[illegible]	0.075$\pm$0.008
[illegible]	0.040$\pm$0.006	0.069$\pm$0.006	0.051$\pm$0.004	0.055$\pm$0.008
[illegible]	[illegible]	[illegible]$\pm$0.033	0.401$\pm$0.001	0.155$\pm$0.015
[illegible]	0.537$\pm$0.006	0.421$\pm$0.011	0.511$\pm$0.007	0.476$\pm$0.013
0.150$\pm$0.017	0.061$\pm$0.003	0.059$\pm$0.008	0.048$\pm$0.001	0.075$\pm$0.014
0.468$\pm$0.020	OOM	OOM	0.458$\pm$0.001	0.372$\pm$0.008
0.351$\pm$0.033	0.219$\pm$0.015	0.123$\pm$0.010	0.226$\pm$0.011	0.144$\pm$0.008
-	0.004	0.004	0.006	0.002

Xu, H., Pang, G., Wang, Y., and Wang, Y. **Deep isolation forest for anomaly detection**. IEEE Transactions on Knowledge and Data Engineering, pp. 1–14, 2023.

Motivation: Challenges

Due to the unsupervised nature of anomaly detection, designing deep anomaly detection models is a journey of ***finding reasonable supervisory signals*** *.*

Current Deep Anomaly Detectors

- **Reconstruction-based Generative methods**

  Employ various kinds of autoencoders, generative adversarial networks, or prediction models

  - *Over-emphasizing low-level details (reducing errors in each fine-grained point)*

- **One-class-based Methods**

  Construct a model (hypersphere or hyperplane) that can describe the data "normality"

  - *One-class assumption might be vulnerable when there are more than one prototype in normal data*

- **Self-supervised Contrastive Methods**

  Define transformation operation to obtain augmented samples and perform proxy tasks

  - *More superior performance, but most transformation operations are not applicable in non-image data*

As for non-perceptual tabular data, it is still a non-trivial task to define suitable supervisory signals to actuate deep learning models.

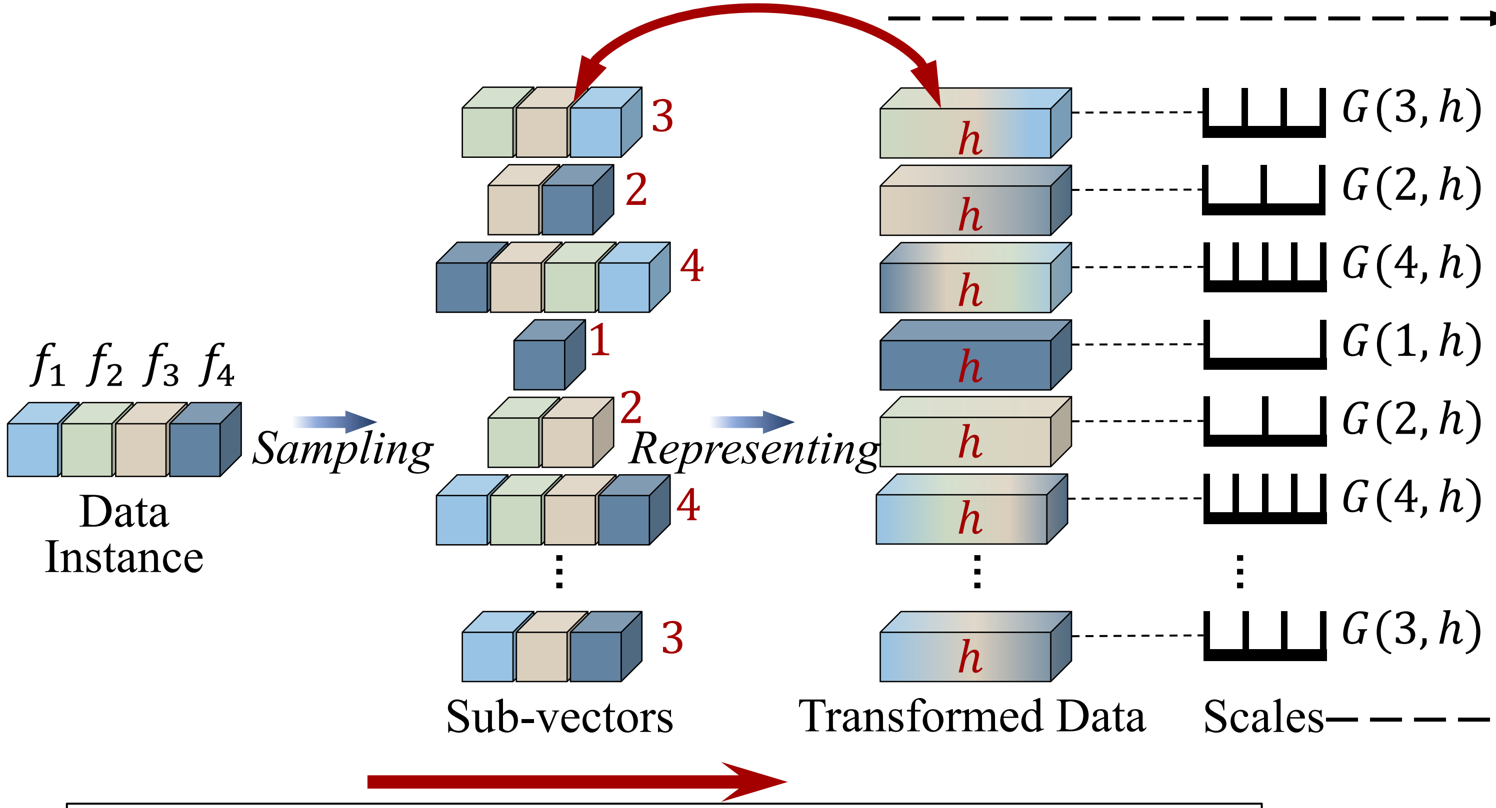
Our Method SLAD: Scale

Definition of "Scale" in real-life

- Scale indicates the ratio between the **real size** of something and **its size on a map, model, or diagram**.

Definition of "Scale" in Tabular data (multi-dimensional vectors)

*Scale is defined as the **mathematical relationship** between **the dimensionality of sub-vectors** and **that of the representations***



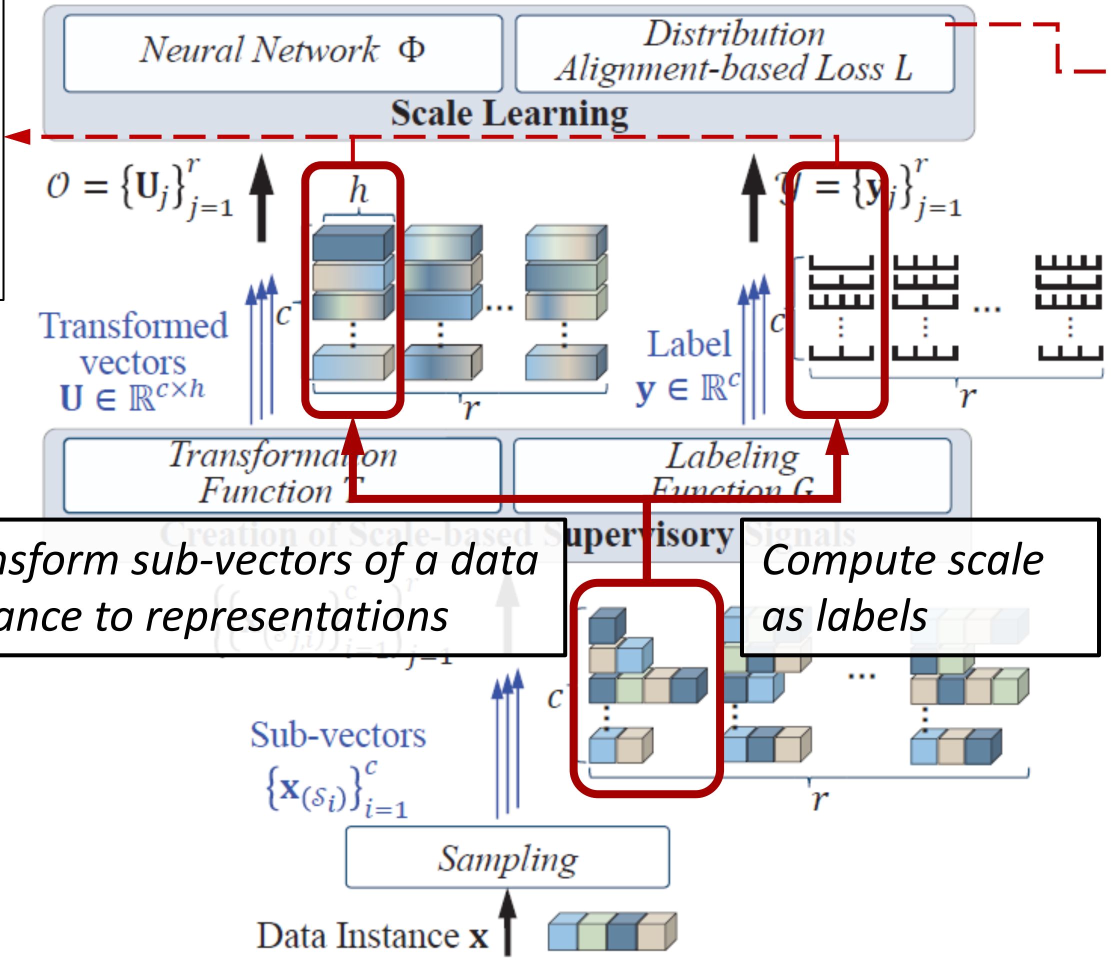
*Scales serve as labels attached to transformed data, supplying **labeled data** for driving neural networks on tabular data*

Sub-vectors of tabular data instances are first transformed to representations via random linear projection

☐ Scale Learning-based deep Anomaly Detection method (SLAD)

Training Data

*Each training sample in scale learning is a group of representations (**data**) and their scales (**label**).*



Transform sub-vectors of a data instance to representations

Compute scale as labels

Training:

Predictions** and **scale labels** are converted to two distributions; Scale learning is defined as a **distribution alignment task.

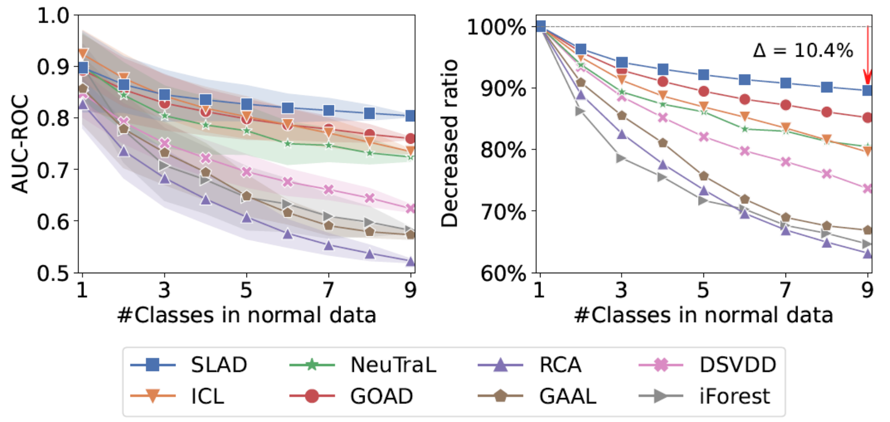
Anomaly detection:

*Through this proxy task, our approach models **inherent regularities and patterns**, which well describes data "normality". Errors computed through the loss function can indicate abnormal degrees of incoming data.*

☐ Empirical Results

- ➢ SLAD Significantly outperforms self-supervised/reconstruction-based/one-class SOTA methods;
- ➢ SLAD can better handle data with complicated normal data (with multiple prototypes);
- ➢ Scale learning has Inlier-priority property;

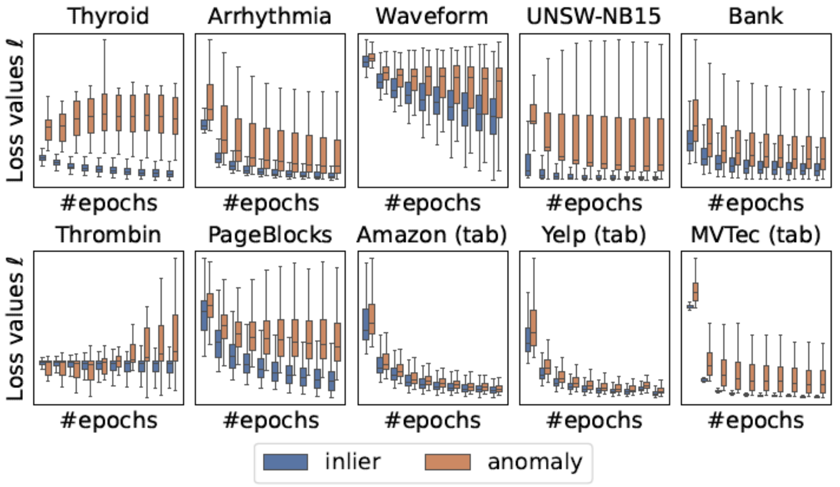
✓ Capability to Handle Complicated Normal Data



✓ Effectiveness

	DATA	SLAD	ICL	NeuTraL	GOAD	RCA	GAAL	DSVDD	iForest
AUC-ROC	Thyroid	**0.995 ± 0.001**	0.974 ± 0.015	0.985 ± 0.002	0.952 ± 0.005	0.934 ± 0.005	0.768 ± 0.096	0.930 ± 0.032	0.988 ± 0.002
	Arrthymia	**0.825 ± 0.007**	0.784 ± 0.048	0.805 ± 0.025	0.806 ± 0.008	0.767 ± 0.009	0.704 ± 0.082	0.807 ± 0.008	0.814 ± 0.007
	Waveform	**0.812 ± 0.047**	0.649 ± 0.048	0.621 ± 0.023	0.604 ± 0.022	0.626 ± 0.019	0.732 ± 0.074	0.516 ± 0.012	0.718 ± 0.019
	UNSW-NB15	**0.941 ± 0.004**	0.918 ± 0.010	0.916 ± 0.017	0.903 ± 0.003	0.935 ± 0.001	0.796 ± 0.060	0.902 ± 0.028	0.758 ± 0.016
	Bank	**0.730 ± 0.004**	0.724 ± 0.014	0.720 ± 0.018	0.587 ± 0.006	0.699 ± 0.003	0.655 ± 0.032	0.608 ± 0.057	0.723 ± 0.008
	Thrombin	**0.939 ± 0.007**	OOM	0.460 ± 0.033	0.839 ± 0.011	0.916 ± 0.000	OOM	0.520 ± 0.046	0.898 ± 0.008
	PageBlocks	**0.972 ± 0.004**	0.909 ± 0.025	0.961 ± 0.002	0.670 ± 0.006	0.864 ± 0.002	0.765 ± 0.032	0.904 ± 0.009	0.927 ± 0.005
	Amazon (tab)	**0.605 ± 0.007**	0.592 ± 0.005	0.570 ± 0.036	0.500 ± 0.000	0.538 ± 0.008	0.495 ± 0.032	0.539 ± 0.013	0.565 ± 0.008
	Yelp (tab)	0.658 ± 0.014	**0.664 ± 0.009**	0.627 ± 0.027	0.501 ± 0.000	0.585 ± 0.008	0.584 ± 0.039	0.593 ± 0.032	0.609 ± 0.007
	MVTec (tab)	**0.812 ± 0.009**	0.778 ± 0.010	0.788 ± 0.009	0.666 ± 0.030	0.663 ± 0.022	0.675 ± 0.026	0.806 ± 0.014	0.757 ± 0.011
AUC-PR	Thyroid	**0.921 ± 0.012**	0.726 ± 0.070	0.824 ± 0.018	0.778 ± 0.008	0.654 ± 0.012	0.429 ± 0.133	0.470 ± 0.030	0.783 ± 0.037
	Arrthymia	0.604 ± 0.006	0.572 ± 0.038	0.589 ± 0.022	0.631 ± 0.005	0.562 ± 0.009	0.505 ± 0.071	0.646 ± 0.008	**0.633 ± 0.021**
	Waveform	**0.432 ± 0.132**	0.123 ± 0.040	0.095 ± 0.014	0.079 ± 0.004	0.088 ± 0.008	0.148 ± 0.060	0.059 ± 0.002	0.111 ± 0.005
	UNSW-NB15	0.858 ± 0.003	**0.859 ± 0.005**	0.811 ± 0.014	0.813 ± 0.005	0.542 ± 0.009	0.470 ± 0.230	0.794 ± 0.028	0.111 ± 0.006
	Bank	**0.470 ± 0.003**	0.468 ± 0.015	0.445 ± 0.018	0.300 ± 0.006	0.423 ± 0.002	0.370 ± 0.050	0.315 ± 0.059	0.449 ± 0.013
	Thrombin	**0.625 ± 0.014**	OOM	0.038 ± 0.002	0.648 ± 0.013	0.587 ± 0.003	OOM	0.074 ± 0.023	0.421 ± 0.017
	PageBlocks	**0.872 ± 0.016**	0.799 ± 0.033	0.871 ± 0.008	0.449 ± 0.010	0.739 ± 0.004	0.500 ± 0.034	0.746 ± 0.017	0.705 ± 0.015
	Amazon (tab)	**0.120 ± 0.002**	0.117 ± 0.001	0.114 ± 0.011	0.095 ± 0.000	0.105 ± 0.003	0.099 ± 0.008	0.107 ± 0.005	0.112 ± 0.002
	Yelp (tab)	**0.153 ± 0.005**	**0.153 ± 0.003**	**0.153 ± 0.015**	0.097 ± 0.000	0.127 ± 0.005	0.125 ± 0.012	0.135 ± 0.013	0.132 ± 0.003
	MVTec (tab)	**0.778 ± 0.009**	0.740 ± 0.009	0.751 ± 0.011	0.606 ± 0.032	0.604 ± 0.022	0.618 ± 0.028	0.771 ± 0.017	0.698 ± 0.011

➢ Inlier-priority Property



Open-source contribution

- The source code is available at https://github.com/xuhongzuo/scale-learning
- **Our method is also included in the DeepOD package.** https://github.com/xuhongzuo/deepod

Python Deep Outlier/Anomaly Detection (DeepOD)

Python Package using Conda passing | Python Package using pip passing | coverage 80% | Downloads 2k

Usages

DeepOD can be used in a few lines of code. This API style is the same with sklearn and PyOD.

```
# unsupervised methods
from deepod.models.dsvdd import DeepSVDD
clf = DeepSVDD()
clf.fit(X_train, y=None)
scores = clf.decision_function(X_test)

# weakly-supervised methods
from deepod.models.devnet import DevNet
clf = DevNet()
clf.fit(X_train, y=semi_y) # semi_y uses 1 for known anomalies,
scores = clf.decision_function(X_test)
```

Easy, consistent, and clear API to use different anomaly detection models

Supported Models

Detection models:

Model	Venue	Year	Type	Title
Deep SVDD	ICML	2018	unsupervised	Deep One-Class Classification
REPEN	KDD	2018	unsupervised	Learning Representations of Ultrahigh-dimensional Data for Random Distance-based Outlier Detection
RDP	IJCAI	2020	unsupervised	Unsupervised Representation Learning by Predicting Random Distances
RCA	IJCAI	2021	unsupervised	RCA: A Deep Collaborative Autoencoder Approach for Anomaly Detection
GOAD	ICLR	2020	unsupervised	Classification-Based Anomaly Detection for General Data
NeuTraL	ICML	2021	unsupervised	Neural Transformation Learning for Deep Anomaly Detection Beyond Images
ICL	ICLR	2022	unsupervised	Anomaly Detection for Tabular Data with Internal Contrastive Learning
DIF	TKDE	2023	unsupervised	Deep Isolation Forest for Anomaly Detection
SLAD	ICML	2023	unsupervised	Fascinating Supervisory Signals and Where to Find Them: Deep Anomaly Detection with Scale Learning
DevNet	KDD	2019	weakly-supervised	Deep Anomaly Detection with Deviation Networks
PReNet	KDD	2023	weakly-	Deep Weakly-supervised Anomaly Detection

13 SOTA Anomaly detection models

We are working on a new feature -- by simply setting a few parameters, different deep anomaly detection models can not only handle different data types.

- We have finished some attempts on partial models like Deep SVDD, DevNet, Deep SAD, PReNet and DIF. These models can use temporal networks like LSTM, GRU, TCN, Conv, Transformer to handle time series data.
- *Future work*: we also want to implement several network structure, so as to processing more data types like graphs and images by simply plugging in corresponding network architecture.

Different network structures (MLP, Transformer, GRU, LSTM, Conv, ...) to handle both tabular and time series data

THANKS

PDL

xuhongzuo13@nudt.edu.cn or hongzuoxu@126.com