Deep Orthogonal Hypersphere Compression for Anomaly Detection

Yunhe Zhang[1,2], Yan Sun[1,3], Jinyu Cai[4], Jicong Fan[1,2]

[1]School of Data Science, The Chinese University of Hong Kong, Shenzhen, China
[2]Shenzhen Research Institute of Big Data, Shenzhen, China
[3]School of Computing, National University of Singapore, Singapore
[4]Institute of Data Science, National University of Singapore, Singapore

Motivation

- Minimizing the sum of squares of the difference between each data point and the center cannot guarantee that the learned decision boundary is a standard hypersphere.
- In high-dimensional space the normal data enclosed by a hypersphere are all far away from the center with high probability (***soap-bubble***). It means that there is no normal data around the center of the hypersphere; whereas anomalous data can still fall into the region.
- The distribution of normal data in the hypersphere is extremely sparse because of the high dimensionality and limited training data. A high sparsity increases the risk of detecting anomalous data as normal.

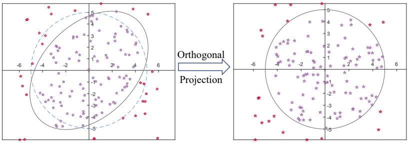


Figure: Toy example of decision boundaries with and without the orthogonal projection layer. Blue circle: assumed decision boundary; black ellipse: actual decision boundary; purple points: normal data; red points: abnormal data.

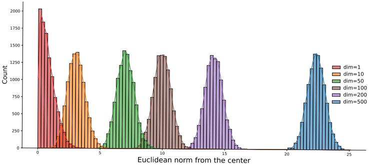


Figure: Soap-bubble phenomenon showed by the histogram of distances from the center of 10^4 samples drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$.

Deep Orthogonal Hypersphere Compression for Anomaly Detection

- Illustration of the proposed Deep Orthogonal Hypersphere Contraction (DOHSC) and Deep Orthogonal Bi-Hypersphere Compression (DO2HSC) methods.

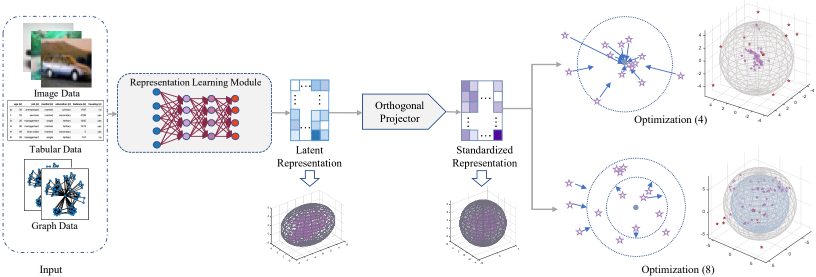


Figure: Architecture of the proposed models *(right top: DOHSC; right bottom: DO2HSC)*. Herein, 2-D visualizations show the trends of training data when applying two optimizations and 3-D visualizations illustrate the detection results obtained by them, respectively.

Hypersphere based Anomaly Detection

We first construct an auto-encoder and utilize the latent representation $\mathbf{Z} = f_{\mathcal{W}}^{\text{enc}}(\mathbf{X})$ to initialize a decision region's center $\boldsymbol{c}$ according to Deep SVDD, i.e, $\boldsymbol{c} = \frac{1}{n}\sum_{i=1}^{n} f_{\mathcal{W}}^{\text{enc}}(\mathbf{x}_i)$. Then the objective function is formulated as:

$$\min_{\mathcal{W}} \frac{1}{n}\sum_{i=1}^{n} \| f_{\mathcal{W}}^{\text{enc}}(\mathbf{x}_i) - \boldsymbol{c}\|^2 + \frac{\lambda}{2}\sum_{l=1}^{L} \|\mathbf{W}_l\|_F^2, \tag{1}$$

where the regularization is to reduce over-fitting.

- The inconsistency between the hypersphere assumption and the actual solution stems from the following two points: **1)** the learned features have different variances and **2)** the learned features are correlated.
- Towards handling these problems, we add the orthogonal projection layer for DOHSC and DO2HSC to pursue orthogonal features of latent representation.

Practical Solution of DOHSC

- Objective Function:

$$\min_{\Theta,\mathcal{W}} \frac{1}{b}\sum_{i=1}^{b} \|\tilde{\boldsymbol{z}}_i - \tilde{\boldsymbol{c}}\|^2 + \frac{\lambda}{2}\sum_{\mathbf{W}\in\mathcal{W}} \|\mathbf{W}\|_F^2, \tag{2}$$

where $\tilde{\boldsymbol{c}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i$ will be **fixed** until optimization is completed, $\tilde{\boldsymbol{z}}$ is the learned orthogonal representation.

- After the training stage, the decision boundary $\hat{r}$ will be **fixed** based on the $1-\nu$ percentile of the training data distance distribution:

$$\hat{r} = \arg\min_{r} \mathcal{P}(\mathbf{D} \leq r) \geq \nu \tag{3}$$

where $\mathbf{D} := \{d_i\}_{i=1}^{N}$ follows a sampled distribution $\mathcal{P}$, and $d_i = \|\tilde{\mathbf{z}}_i - \tilde{\mathbf{c}}\|$.

- Accordingly, the anomalous score of i-th instance is defined as follows:

$$s_i = d_i^2 - \hat{r}^2, \tag{4}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_n)$.

Practical Solution of DO2HSC

- To achieve the contraction target of the bi-hypersphere, the pretraining stage (i.e., performing DOHSC first) is necessary to determine its decision boundary ($r_{\min}$ and $r_{\max}$).

$$r_{\max} = \arg\min_{r} \mathcal{P}(\mathbf{D} \leq r) \geq \nu, \quad r_{\min} = \arg\min_{r} \mathcal{P}(\mathbf{D} \leq r) \geq 1 - \nu. \tag{5}$$

- Then the objective function becomes:

$$\min_{\Theta, \mathcal{W}} \frac{1}{b} \sum_{i=1}^{b} \left(\max\{d_i, r_{\max}\} - \min\{d_i, r_{\min}\}\right) + \frac{\lambda}{2} \sum_{\mathbf{W} \in \mathcal{W}} \|\mathbf{W}\|_F^2. \tag{6}$$

- Accordingly, the anomalous score of i-th instance is defined as follows:

$$s_i = (d_i - r_{\max}) \cdot (d_i - r_{\min}), \tag{7}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_n)$.

Numerical Results

Table: Average AUCs (%) in one-class anomaly detection on CIFAR-10.

Normal Class	Airplane	Auto Mobile	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
Deep SVDD	61.7	65.9	50.8	59.1	60.9	65.7	67.7	67.3	75.9	73.1
OCGAN	75.7	53.1	64.0	62.0	72.3	62.0	72.3	57.5	82.0	55.4
DROCC*	**82.1**	64.8	69.2	64.4	72.8	66.5	68.6	67.5	79.3	60.6
HRN-L2	80.6	48.2	64.9	57.4	**73.3**	61.0	74.1	55.5	79.9	71.6
HRN	77.3	69.9	60.6	64.4	71.5	67.4	77.4	64.9	82.5	77.3
PLAD	**82.5**	80.8	68.8	65.2	71.6	71.2	76.4	73.5	80.6	80.5
DOHSC	80.3	**81.0**	**70.4**	**68.0**	72.1	**72.4**	**83.1**	**74.1**	**83.3**	**81.1**
	(0.0)	**(0.0)**	**(1.9)**	**(1.8)**	(0.0)	**(2.1)**	**(0.0)**	**(0.4)**	**(0.7)**	**(0.7)**
DO2HSC	81.3	**82.7**	**71.3**	**71.2**	**72.9**	**72.8**	**83.0**	**75.5**	**84.4**	**82.0**
	(0.2)	**(0.3)**	**(0.4)**	**(1.3)**	**(2.1)**	**(0.2)**	**(0.6)**	**(0.4)**	**(0.5)**	**(0.9)**

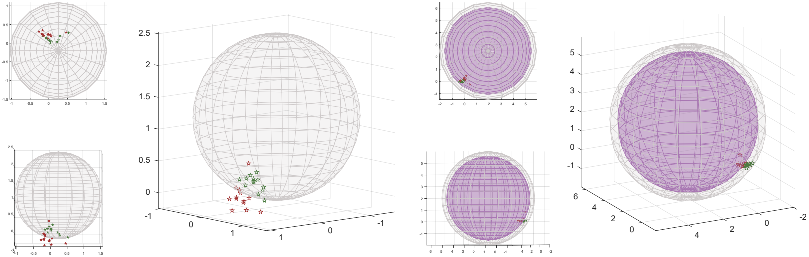
Table: Average F1-scores on tabular datasets.

	Thyroid	Arrhythmia
OCSVM	0.56 ± 0.01	0.64 ± 0.01
Deep SVDD	0.73 ± 0.00	0.54 ± 0.01
LOF	0.54 ± 0.01	0.51 ± 0.01
GOAD	0.75 ± 0.01	0.52 ± 0.02
DROCC	0.78 ± 0.03	0.69 ± 0.02
PLAD	0.77 ± 0.01	**0.71 ± 0.02**
DOHSC	**0.92 ± 0.01**	0.70 ± 0.03
DO2HSC	**0.98 ± 0.59**	**0.74 ± 0.02**

Table: Average AUCs for graph-level anomaly detection algorithms.

	COLLAB			MUTAG		ER_MD	
	0	1	2	0	1	0	1
SP+OCSVM	0.5910 ± 0.0000	0.8397 ± 0.0000	0.7902 ± 0.0000	0.5917 ± 0.0000	0.2608 ± 0.0000	0.4092 ± 0.0000	0.3824 ± 0.0000
WL+OCSVM	0.5122 ± 0.0000	0.8054 ± 0.0000	0.7996 ± 0.0000	0.6509 ± 0.0000	0.2960 ± 0.0000	0.4571 ± 0.0000	0.3262 ± 0.0000
NH+OCSVM	0.5976 ± 0.0000	0.8054 ± 0.0000	0.6414 ± 0.0000	0.7959 ± 0.0274	0.1679 ± 0.0062	0.5155 ± 0.0200	0.3648 ± 0.0000
RW+OCSVM	–	–	–	0.8698 ± 0.0000	0.1504 ± 0.0000	0.4820 ± 0.0000	0.3484 ± 0.0000
OCGIN	0.4217 ± 0.0606	0.7565 ± 0.2035	0.1906 ± 0.0857	0.8491 ± 0.0424	0.7466 ± 0.0168	0.5645 ± 0.0323	0.4358 ± 0.0538
infoGraph+DSVDD	0.5662 ± 0.0597	0.7926 ± 0.0986	0.4062 ± 0.0978	0.8805 ± 0.0448	0.6166 ± 0.2052	0.5312 ± 0.1545	0.5082 ± 0.0704
GLocalKD	0.4638 ± 0.0003	0.4330 ± 0.0016	0.4792 ± 0.0004	0.3952 ± 0.2258	0.2965 ± 0.2641	0.5781 ± 0.1790	**0.7154 ± 0.0000**
OCGTL	0.6504 ± 0.0433	0.8908 ± 0.0239	0.4029 ± 0.0541	0.6570 ± 0.0210	0.7579 ± 0.2212	0.2755 ± 0.0317	0.6915 ± 0.0207
DOHSC	**0.9185 ± 0.0455**	**0.9755 ± 0.0030**	**0.8826 ± 0.0250**	**0.8822 ± 0.0432**	**0.8115 ± 0.0279**	**0.6620 ± 0.0308**	0.5184 ± 0.0793
DO2HSC	**0.9390 ± 0.0025**	**0.9836 ± 0.0115**	**0.8835 ± 0.0118**	**0.9089 ± 0.0609**	**0.8250 ± 0.0790**	**0.6867 ± 0.0226**	**0.7351 ± 0.0159**

Visualization Results



(a) DOHSC

(b) DO2HSC

Figure: Anomaly detection comparison between DOHSC and DO2HSC on MUTAG.