Appendix of "Multi-view Deep Anomaly Detection: A Systematic Exploration"

Siqi Wang, Jiyuan Liu, Guang Yu, Xinwang Liu, Sihang Zhou, En Zhu,
Yuexiang Yang, Jianping Yin, Wenjing Yang

I. Classic AD methods

As a classic problem, AD has constantly received attention from the machine learning community. Classic solutions to AD usually fall into the following categories: 1) *Boundary based methods*. Such methods learn to establish a compact decision boundary to enclose training data from the given normal class, while this boundary is then used to separate the normal and abnormal class. Two most representative methods are one-class support vector machine (OCSVM) [1] and support vector data description (SVDD) [2]. 2) *Density based methods*. Methods like one-class Gaussian Mixture Model (OCGMM) [3] and one-class Parzen density estimation [4] usually estimate the data density distribution by given training data of the normal class, and detect data at the sparse region as anomalies. 3) *Reconstruction based methods*. Such methods assume that data from normal class can be well reconstructed from its low-dimensional embedding, while poor reconstruction would be observed on the untrained abnormal class. The reconstruction can be performed by methods like principal component analysis (PCA) [5] or shallow auto-encoder network (AE) [6]. In addition to the above three categories, some creative solutions are also proposed to solve AD problem. To name a few, Juszczak et al. [7] propose to build minimal spanning tree (MST) to model the normal class. Angiulli [8] introduces a series of prototypes to describe the domain of normal class. Désir et al. [9] customize the random forest model to AD and propose one-class random forest. Classic AD methods are often intuitive, yet they typically handle with relatively simple tabular data from a single view.

II. Classic Multi-view Learning Methods

Multi-view learning has been recognized as a vital realm for a long period. It explores complementary clues of multiple data views, so as to boost the performance of a certain task. Many regular tasks have already been discussed in the context of the multi-view case, e.g. multi-view classification [10], multi-view clustering [11]) and multi-view feature selection [12]. Those tasks then give rise to many multi-view learning models, such as multi-view support vector machine [13], multi-view subspace clustering [14], multiple kernel k-means [15], etc. A comprehensive review of classic multi-view learning can be found in [16]. In many works, multi-view learning tasks are often formulated as a convex or non-convex optimization problem that can be solved by certain classic optimization strategies. However, many of such solutions are of high computational complexity, making it hard to scale them to those large-scale tasks. Besides, compared with deep learning based methods, they also require high-quality features extraction in advance, which is the foundation of their effectiveness and efficiency.

Manuscript received June 4, 2021.

III. Other Multi-view Deep AD Baselines

In addition to 11 baseline solutions that have been presented in the manuscript, we also design 4 additional baseline solutions, as introduced below. The evaluation results of them are reported in Table III.

A. Deep Belief Networks based Solution

Apart from those fusion functions given in the manuscript, multi-view fusion can also be implemented by the classic restricted boltzmann machines (RBM) from the energy view [17]. To be more specific, a RBM is added to the top of all encoders to perform fusion [18]. With the concatenation of embeddings of all views $\mathbf{v} = Cat(\{\mathbf{h}^{(v)}\}_{v=1}^{V})$, it serves as the input to the visible layer of the RBM. To learn the joint embedding $\mathbf{h}$ given by RBM's hidden layer, it is required to minimize the following energy function:

$$E(\mathbf{v},\mathbf{h}) = -\mathbf{v}^\top \cdot \mathbf{W} \cdot \mathbf{h} - \mathbf{b}^\top \cdot \mathbf{v} - \mathbf{a}^\top \cdot \mathbf{h} \tag{1}$$

where $\mathbf{W}, \mathbf{a}, \mathbf{b}$ are learnable parameters of the RBM. A joint probability of $\mathbf{v}$ and $\mathbf{h}$ can be calculated by:

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{C}\exp(-E(\mathbf{v},\mathbf{h})) \tag{2}$$

where C is the partition function for normalization. The maximization of joint probability (the minimization of energy) can be realized by gradient descent and contrastive divergence. After the minimization, the joint embedding $\mathbf{h}$ and the reconstruction of $\mathbf{v}$ can both be obtained by sampling. In addition to the fusion function, the encoders and decoders of the model can also be implemented by RBMs. In this way, we can stack the RBMs into a deep belief network (DBN). However, DBN is not specifically designed for more complex data with certain structures (e.g. 2D images), which restricts its application to many practical scenarios.

B. Generalized DCCA based Solution

It is easy to discover that alignment functions given in the manuscript require to compute the alignment for each view pair, and then add their results up to obtain the final alignment measurement. By contrast, a deep generalized canonical correlation analysis (DGCCA) [19] is proposed to naturally adapt to the alignment of any number of views. The key to generalized correlation based alignment is to encourage embeddings from all views to be closed to a common representation $\mathbf{G}$. Given the linear mapping parameterized by $\mathbf{U}^{(v)}$ and the embedding matrix $\mathbf{H}^{(v)}$ for the v_{th} view, it aims to sovle the following optimization problem:

$$\min_{\mathbf{G},\mathbf{U}^{(v)}} \frac{1}{V}\sum_{v=1}^{V} \|\mathbf{G}-\mathbf{U}^{(v)}\cdot\mathbf{H}^{(v)}\|_F^2, \quad s.t.\ \mathbf{G}^\top\mathbf{G}=\mathbf{I}_r, \tag{3}$$

in which $\mathbf{I}_r$ is the identity matrix with size r. The details to solve the goal in Eq. 3 is shown in [19]. In our experiments, the performance of DGCCA is shown to be similar to DCCA, so we leave it to supplementary material.

C. Soft-boundary DSVDD based Solution

By contrast, soft-boundary DSVDD takes the noises in training data into account, and introduces a soft boundary to allow noisy data not to be mapped into the hyper-sphere. It solves the optimization problem below:

$$\begin{aligned}\min_{\boldsymbol{R},\boldsymbol{\theta}^{(v)}} \quad & R^2+\frac{1}{\nu N}\sum_{n=1}^{N}\max\{0,||Enc^{(v)}(\mathbf{x}_n^{(v)}))-\mathbf{c}^{(v)}||_2^2-R^2\} \\ & +\frac{\lambda}{2}||\boldsymbol{\theta}_E^{(v)}||_F^2,\end{aligned} \tag{4}$$

where R represents the radius of hyper-sphere, and $\nu > 0$ controls the softness of the boundary. Likewise, the above optimization problem can also be solved by gradient descent. Soft-boundary DSVDD can be tailored for multi-view deep AD by the same way of simplified DSVDD, but it typically performs worse than the simplified version.

D. Discriminative Self-supervision based Solution

Discriminative pretext task based self-supervised learning is an newly-emerging technique [20]. Instead of conducting generation, discriminative pretext tasks require to perform representation learning by classification, because discriminative DNNs are usually considered to be more powerful than generative DNNs in representation learning. As AD only has training data from a single class, multiple pseudo classes need to be created first to enable classification, which is the key to discriminative pretext tasks. Specifically, we define a transformation set $\mathcal{T}=\{T_1,T_2,\cdots,T_m\}$. For each transformation $T_i\in\mathcal{T}$, it can transform a multi-view datum $\{\mathbf{x}_n^{(v)}\}_{v=1}^V$ into a new single-view datum $\tilde{\mathbf{x}}_{i,n}=T_i(\{\mathbf{x}_n^{(v)}\}_{v=1}^V)$. After the transformation, all data yielded by the transformation T_i are collected as the i_{th} pseudo class $\mathcal{C}_i'=\{\tilde{\mathbf{x}}_{i,n}\}_{n=1}^N$. In this way, a set of m pseudo classes can be collected as $\{\mathcal{C}_i'\}_{i=1}^m$. Subsequently, those pseudo classes can be used to train a discriminative DNNs by standard cross-entropy loss. As to inference, a multi-view datum $\{\mathbf{x}_{test}^{(v)}\}_{v=1}^V$ is first transformed into m new single-view datum $\{\tilde{\mathbf{x}}_{i,test}\}_{i=1}^m$, Golan et al. [20] show that a simple way to obtain the score of $\mathbf{x}_{test}^{(v)}$ can be:

$$\mathcal{S}(\{\mathbf{x}_{test}^{(v)}\}_{v=1}^V)=\frac{1}{m}\sum_{i=1}^{m}p(i|\tilde{\mathbf{x}}_{i,test}) \tag{5}$$

where $p(i|\tilde{\mathbf{x}}_{i,test})$ refers to the confidence that $\tilde{\mathbf{x}}_{i,test}$ belongs to the i_{th} pseudo class. When it comes to the design of T_i, we are inspired by [21] and adopt a simple but generic method: we first map the input data of each view to D-dimensional embeddings by random projection, then embeddings of different views are permuted by a certain order and concatenated into a joint embedding. By varying the parameters of random projection or the permutation order, we can obtain multiple transformations in $\mathcal{T}$. In this way, we can not only create sufficient transformations to create pseudo classes, but also exploit the unique feature of multi-view data. However, although discriminative self-supervised learning has achieved remarkable success in other realms, our empirical evaluations show that it is usually inferior to generative self-supervised learning in multi-view deep AD.

IV. Details of Multi-view Datasets

In Table I and II, we showcase the detailed information of all multi-view benchmark datasets used in this paper, including the total number of samples, number of views and the total number of classes.

TABLE I: Details of multi-view deep one-class classification datasets. *MedMNIST* and *MvTecAD* are introduced in Table II, since they are composed of multiple subsets.

Dataset		Number of Samples	Number of Views	Number of Classes
Classic	BBC	2012	2	5
	BDGP	2500	3	5
	Caltech20	2386	6	20
	Citeseer	3312	4	6
	Cora	2708	4	7
	Reuters	7200	5	6
	Wiki	2866	2	10
	AwA	30475	6	50
	NUS-Wide	23953	5	31
	SunRGBD	10335	2	45
	YtFace	101499	11	31
Image	MNIST	70000	6	10
	FashionMNIST	70000	6	10
	Cifar10	60000	6	10
	Cifar100	60000	6	10
	SVHN	99289	6	10
	Cat_vs_Dog	24931	4	2
	CMU-MOSE	23500	3	7
	DriverAD	453750	4	2
	MedMNIST	-	-	-
	MvTecAD	-	-	-
Video	UCSDped1	143259	2	2
	UCSDped2	64061	2	2
	UMN_scene1	19648	2	2
	UMN_scene2	53561	2	2
	UMN_scene3	34294	2	2
	Avenue	219026	2	2
	ShanghaiTech	1419412	2	2

V. Implementation Details

For tabular input data, we leverage a fully-connected DNN to encode them into latent embeddings. The fully-connected

TABLE II: Details of *MedMNIST* and *MvTecAD* datasets.

Dataset		Number of Samples	Number of Views	Number of Classes
MedMNIST	Path	107180	6	9
	Derma	10015	6	7
	OCT	109309	6	4
	Pneumonia	5856	6	2
	Retina	1600	6	5
	Breast	780	6	2
	Axial	58850	6	11
	Coronal	25221	6	11
	Sagittal	25221	6	11
MvTecAD	Bottle	292	4	2
	Cable	374	4	2
	Capsule	351	4	2
	Carpet	397	4	2
	Grid	342	4	2
	Hazelnut	501	4	2
	Leather	369	4	2
	Metal Nut	335	4	2
	Pill	434	4	2
	Screw	480	4	2
	Tile	347	4	2
	Toothbrush	102	4	2
	Transistor	313	4	2
	Wood	326	4	2
	Zipper	391	4	2

DNN has $512-128-32$ hidden layers, which are equipped with batch-normalization (bn) and ReLu activation function ($relu$). For 2-D input data on video based multi-view datasets, we implement the encoders by a convolutional neural networks with the following architecture: $conv(3,2)-bn-relu-conv(3,2)-bn-relu-conv(3,2)-bn-relu-reshape-fc(2048,32)$, where $conv(3,2)$ denotes the 2-D convolution operation with kernel size 3 and stride 2 and fc denotes a fully-connected layer. As to decoders, we simply adopt a symmetric DNN architecture to realize decoding, while the decoders for 2-D input data are implemented by deconvolution operation. As to training, since the fine-tuning of hyperparameters for AD is difficult, we empirically set the training epochs of video based multi-view datasets to be 5 or 10, while the rest of datasets are set to be 20. The batch size is typically selected from 16, 32, 64 and 128, according to the size of training set. For each dataset, the batch size is fixed for all baseline solutions. Meanwhile, the default Adam optimizer in PyTorch toolbox[1] is used. The weight of alignment loss is set to $\alpha = 0.1$. The weight of L_2-norm regularization is set to 0. For tensor fusion, the rank R is set to 8. For energy based fusion, we leverage a deep belief network that share the same hidden layers with the aforementioned fully-connected encoder. For similarity based alignment, we simply set the margin to be 0. For generative self-supervision based methods, we adopt a fully-connected neural network to perform fusion. For the discriminative self-supervised method, we use 16, 3, 4, 1 random projections for data with 2-view, 3-view, 4-view and more than 4 views respectively, which result in $16\times 2! = 32$, $6\times 3! = 36$, $2\times 4! = 48$ and $1\times V! = V!$ pseudo classes for four cases. The classifier shares the same hidden layer architecture with the previous fully-connected encoder network, and a fully-connected layer and a softmax layer are added to its top for classification.

[1]https://pytorch.org/

VI. Additional Experimental Results

Table III reports the performance comparison on selected existing multi-view datasets between four miscellaneous baselines in this supplementary material and eleven ones in the manuscript (DBN, DGCCA, DSV-B, CLAS denote the baselines introduced in Sec. III-A-III-D respectively). Meanwhile, Table IV, V and VI present the AUROC results on the subsets of *MedMNIST* and *MvTecAD*. Additionally, we show the performance of different baselines and late fusion strategies under AUPR and TNR@95%TPR metrics over all datasets in Table VII - XVIII.

References

[1] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.

[2] D. M. Tax and R. P. Duin, "Support vector data description," *Machine learning*, vol. 54, no. 1, pp. 45–66, 2004.

[3] S. Roberts and L. Tarassenko, "A probabilistic resource allocating network for novelty detection," *Neural Computation*, vol. 6, no. 2, pp. 270–284, 1994.

[4] G. Cohen, H. Sax, A. Geissbuhler *et al.*, "Novelty detection using one-class parzen density estimator. an application to surveillance of nosocomial infections," in *Mie*, 2008, pp. 21–26.

[5] H. Hoffmann, "Kernel pca for novelty detection," *Pattern recognition*, vol. 40, no. 3, pp. 863–874, 2007.

[6] N. Japkowicz, C. Myers, and M. Gluck, "A novelty detection approach to classification," in *International Joint Conference on Artificial Intelligence*, 1995, pp. 518–523.

[7] P. Juszczak, D. M. Tax, E. Pe, R. P. Duin *et al.*, "Minimum spanning tree based one-class classifier," *Neurocomputing*, vol. 72, no. 7-9, pp. 1859–1869, 2009.

[8] F. Angiulli, "Prototype-based domain description for one-class classification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 6, pp. 1131–1144, 2012.

[9] C. Désir, S. Bernard, C. Petitjean, and L. Heutte, "One class random forests," *Pattern Recognition*, vol. 46, no. 12, pp. 3490–3506, 2013.

[10] X. Liu, L. Wang, J. Yin, E. Zhu, and J. Zhang, "An efficient approach to integrating radius information into multiple kernel learning," *IEEE transactions on cybernetics*, vol. 43, no. 2, pp. 557–569, 2013.

[11] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao, "Late fusion incomplete multi-view clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 10, pp. 2410–2423, 2018.

[12] C. Tang, X. Zheng, X. Liu, W. Zhang, J. Zhang, J. Xiong, and L. Wang, "Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[13] L. Houthuys, R. Langone, and J. A. K. Suykens, "Multi-view least squares support vector machines classification," *Neurocomputing*, vol. 282, pp. 78–88, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.12.029

[14] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, 2020. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2877660

[15] J. Liu, X. Liu, J. Xiong, Q. Liao, S. Zhou, S. Wang, and Y. Yang, "Optimal neighborhood multiple kernel clustering with adaptive local kernels," *IEEE Transactions on Knowledge and Data Engineering*, 2020. [Online]. Available: https://doi.org/10.1109/TKDE.2020.3014104

[16] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Information Fusion*, vol. 38, pp. 43–54, 2017.

[17] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[18] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep boltzmann machines." *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 2949–2980, 2014.

[19] A. Benton, H. Khayrallah, B. Gujral, D. A. Reisinger, S. Zhang, and R. Arora, "Deep generalized canonical correlation analysis," in *4th Workshop on Representation Learning for NLP*, 2019, pp. 1–6.

TABLE III: AUROC (%) comparison on selected existing multi-view datasets between thirteen baselines in the manuscript and four miscellaneous ones, including deep belief networks based solution (DBN), generalized DCCA based solution (DGCCA), Soft-boundary DSVDD based solution (DSV-B) and discriminative self-supervision based solution (CLAS).

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
SUM	$94.35_{\pm0.54}$	$81.27_{\pm0.77}$	$99.76_{\pm0.11}$	$83.85_{\pm0.33}$	$87.79_{\pm0.53}$	$65.05_{\pm0.43}$	$88.84_{\pm0.80}$	$63.15_{\pm0.72}$	$67.94_{\pm0.54}$	$84.81_{\pm0.45}$
MAX	$94.35_{\pm0.54}$	$81.36_{\pm0.84}$	$99.69_{\pm0.17}$	$83.86_{\pm0.33}$	$87.79_{\pm0.51}$	$65.04_{\pm0.42}$	$88.93_{\pm0.64}$	$63.34_{\pm0.77}$	$68.25_{\pm0.49}$	$84.63_{\pm0.51}$
NN	$94.35_{\pm0.54}$	$80.98_{\pm0.84}$	$99.77_{\pm0.11}$	$83.87_{\pm0.34}$	$87.78_{\pm0.52}$	$65.03_{\pm0.42}$	$88.85_{\pm0.51}$	$63.27_{\pm0.71}$	$68.56_{\pm0.47}$	$84.55_{\pm0.42}$
TF	$94.35_{\pm0.54}$	$81.03_{\pm0.86}$	$97.79_{\pm1.30}$	$83.87_{\pm0.32}$	$87.78_{\pm0.52}$	$65.05_{\pm0.42}$	$89.24_{\pm1.03}$	$62.76_{\pm0.72}$	$67.24_{\pm0.56}$	$84.37_{\pm0.56}$
DIS	$94.35_{\pm0.54}$	$82.03_{\pm0.80}$	$99.82_{\pm0.08}$	$83.86_{\pm0.34}$	$87.79_{\pm0.53}$	$65.05_{\pm0.42}$	$86.58_{\pm0.77}$	$62.96_{\pm0.67}$	$66.91_{\pm0.64}$	$84.16_{\pm0.43}$
SIM	$94.35_{\pm0.54}$	$81.85_{\pm0.80}$	$99.77_{\pm0.12}$	$83.87_{\pm0.32}$	$87.78_{\pm0.53}$	$65.08_{\pm0.42}$	$86.11_{\pm0.77}$	$62.67_{\pm0.65}$	$67.02_{\pm0.42}$	$84.27_{\pm0.59}$
DCCA	$94.35_{\pm0.54}$	$81.74_{\pm0.84}$	$99.74_{\pm0.14}$	$83.86_{\pm0.34}$	$87.78_{\pm0.52}$	$65.08_{\pm0.42}$	$87.49_{\pm0.84}$	$62.76_{\pm0.67}$	$66.89_{\pm0.48}$	$84.00_{\pm0.49}$
DAE	$94.35_{\pm0.54}$	$81.99_{\pm0.79}$	$99.80_{\pm0.11}$	$83.86_{\pm0.32}$	$87.79_{\pm0.52}$	$65.05_{\pm0.42}$	$85.87_{\pm0.56}$	$62.84_{\pm0.70}$	$66.59_{\pm0.57}$	$84.18_{\pm0.43}$
SVDD	$93.64_{\pm0.59}$	$76.09_{\pm1.51}$	$98.11_{\pm0.24}$	$72.86_{\pm0.65}$	$82.66_{\pm1.08}$	$64.53_{\pm0.40}$	$84.81_{\pm0.54}$	$61.96_{\pm0.47}$	$66.33_{\pm0.78}$	$68.33_{\pm1.22}$
PPRD	$94.35_{\pm0.54}$	$81.13_{\pm1.00}$	$99.55_{\pm0.20}$	$83.86_{\pm0.33}$	$87.78_{\pm0.51}$	$65.03_{\pm0.42}$	$90.93_{\pm0.53}$	$63.51_{\pm0.79}$	$67.71_{\pm0.40}$	$83.39_{\pm0.40}$
SPRD	$94.35_{\pm0.54}$	$79.50_{\pm0.93}$	$99.61_{\pm0.19}$	$83.87_{\pm0.33}$	$87.78_{\pm0.52}$	$65.01_{\pm0.42}$	$90.82_{\pm0.63}$	$63.50_{\pm0.64}$	$68.62_{\pm0.55}$	$84.81_{\pm0.44}$
MODDIS	$94.35_{\pm0.54}$	$59.00_{\pm1.21}$	$78.77_{\pm1.73}$	$78.37_{\pm0.52}$	$86.71_{\pm0.40}$	$64.38_{\pm0.42}$	$86.40_{\pm1.21}$	$59.42_{\pm0.65}$	$63.45_{\pm0.53}$	$46.79_{\pm1.16}$
CAAE	$93.07_{\pm0.50}$	$76.00_{\pm1.34}$	$99.29_{\pm0.16}$	$74.95_{\pm0.45}$	$84.45_{\pm0.57}$	$64.52_{\pm0.59}$	$87.47_{\pm0.56}$	$62.24_{\pm0.63}$	$67.78_{\pm0.69}$	$73.46_{\pm0.86}$
DBN	$95.26_{\pm0.50}$	$63.13_{\pm0.64}$	$93.81_{\pm0.62}$	$84.19_{\pm0.28}$	$87.87_{\pm0.51}$	$65.85_{\pm0.36}$	$84.12_{\pm0.85}$	$63.22_{\pm0.63}$	$65.25_{\pm0.46}$	$79.85_{\pm0.36}$
DGCCA	$94.35_{\pm0.54}$	$81.31_{\pm0.76}$	$99.77_{\pm0.14}$	$83.86_{\pm0.33}$	$87.78_{\pm0.53}$	$65.04_{\pm0.42}$	$86.56_{\pm0.98}$	$63.00_{\pm0.72}$	$67.55_{\pm0.61}$	$83.49_{\pm0.67}$
DSV-B	$93.66_{\pm0.60}$	$76.11_{\pm1.46}$	$98.46_{\pm0.29}$	$73.18_{\pm0.59}$	$83.38_{\pm1.02}$	$64.52_{\pm0.40}$	$84.31_{\pm0.65}$	$62.17_{\pm0.49}$	$66.90_{\pm0.90}$	$68.03_{\pm1.94}$
CLAS	$93.96_{\pm0.53}$	$78.59_{\pm1.33}$	$87.80_{\pm2.46}$	$57.83_{\pm1.14}$	$74.89_{\pm1.30}$	$63.27_{\pm0.48}$	$68.13_{\pm1.43}$	-	$54.44_{\pm0.38}$	$58.13_{\pm0.61}$

TABLE IV: AUROC (%) comparison on the subsets of *MedMNIST*.

Type		Path	Derma	OCT	Pneumonia	Retina	Breast	Axial	Coronal	Sagittal
Fusion	SUM	83.00	70.29	59.39	76.13	65.24	72.67	94.29	95.17	91.51
	MAX	82.69	70.32	60.14	75.45	65.37	72.71	94.28	95.16	91.50
	NN	82.82	70.27	59.24	77.27	64.96	72.57	94.29	95.17	91.48
	TF	82.62	70.13	59.43	75.31	64.90	72.40	94.32	95.16	91.47
Alignment	DIS	83.85	70.21	58.68	78.63	64.85	72.71	94.21	95.16	91.43
	SIM	84.31	70.28	58.95	76.17	64.89	72.87	94.28	95.16	91.45
	DCCA	83.93	69.99	60.78	75.90	64.79	72.69	94.36	95.15	91.41
Tailored	DAE	82.41	66.53	60.77	75.68	61.92	72.90	93.20	94.34	90.41
	SVDD	75.66	59.30	55.61	72.97	58.91	55.85	68.52	58.78	71.47
Self-supervision	PPRD	85.16	70.08	62.33	77.83	65.12	72.72	94.26	95.17	91.47
	SPRD	87.55	70.49	63.23	76.26	65.02	72.66	94.36	95.17	91.52
MDOD	MODDIS	78.52	67.20	50.39	72.74	63.49	71.04	91.22	94.43	90.39
	CAAE	82.50	67.21	61.86	75.40	63.59	72.79	93.34	94.68	90.83

TABLE V: AUROC (%) comparison on partial subsets of *MvTecAD*.

Type		Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal Nut
Fusion	SUM	97.30	91.87	91.82	94.18	63.24	98.00	99.76	88.66
	MAX	97.30	92.17	92.58	94.34	62.16	97.75	99.66	87.05
	NN	97.30	92.26	88.83	94.34	61.15	96.71	99.56	87.68
	TF	97.30	90.89	89.19	94.22	61.15	97.32	99.66	87.93
Alignment	DIS	97.30	91.51	88.35	94.18	62.91	96.54	99.80	89.00
	SIM	97.30	91.85	90.15	94.22	62.57	97.00	99.76	88.51
	DCCA	97.30	91.87	90.19	94.30	62.82	97.29	99.80	89.15
Tailored	DAE	97.22	84.15	69.80	93.42	58.81	81.43	99.08	66.96
	SVDD	94.92	63.87	59.15	93.98	64.16	72.36	97.59	59.38
Self-supervision	PPRD	97.30	92.93	91.54	94.02	62.82	97.11	99.69	88.81
	SPRD	97.30	92.28	91.26	94.18	59.23	97.57	99.69	89.10
MDOD	MODDIS	96.98	73.84	64.74	93.86	59.40	67.07	98.06	43.89
	CAAE	97.06	85.16	76.75	95.02	58.15	93.25	99.32	71.65

[20] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Advances in Neural Information Processing Systems*, 2018, pp. 9758–9769.

[21] L. Bergman and Y. Hoshen, "Classification-based anomaly detection for general data," in *International Conference on Learning Representations*, 2019.

TABLE VI: AUROC (%) comparison on partial subsets of *MvTecAD*.

Type		Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper
Fusion	SUM	82.19	68.15	99.64	91.39	93.46	91.58	96.17
	MAX	81.29	69.17	99.57	91.39	93.00	91.58	95.25
	NN	82.87	67.21	99.60	91.11	92.25	91.40	96.03
	TF	83.33	68.37	98.88	89.72	91.38	91.49	95.09
Alignment	DIS	83.14	66.84	99.68	90.28	91.79	91.58	96.32
	SIM	83.61	67.43	99.64	91.39	92.79	91.58	95.96
	DCCA	83.63	71.49	99.64	90.00	92.21	91.58	95.88
Tailored	DAE	68.99	54.09	97.04	78.33	80.96	89.47	86.66
	SVDD	60.47	50.83	93.36	21.94	69.92	90.88	81.70
Self-supervision	PPRD	80.17	71.33	99.64	90.56	92.50	91.58	92.99
	SPRD	81.48	71.35	99.64	90.83	92.25	91.58	94.75
MDOD	MODDIS	62.58	50.13	94.59	65.83	81.12	92.02	81.12
	CAAE	79.00	56.63	99.60	77.22	88.37	91.14	89.71

TABLE VII: AUPR-normal (%) of different baselines on image based multi-view datasets.

Type		MNIST	FashionMNIST	Cifar10	Cifar100	SVHN	Cat_vs_Dog	MedMNIST	MvTecAD
Fusion	SUM	91.58	69.39	39.38	24.57	41.04	96.56	59.11	78.04
	MAX	91.39	68.87	38.55	23.90	40.74	96.57	59.07	76.98
	NN	91.87	69.68	37.18	23.95	41.17	96.72	59.08	75.84
	TF	91.58	68.05	38.48	23.64	42.27	96.34	58.75	74.96
Alignment	DIS	91.25	69.24	38.88	24.09	42.70	96.83	59.07	76.92
	SIM	91.54	68.96	39.35	24.56	43.44	96.78	59.15	77.17
	DCCA	91.67	66.88	31.88	19.96	43.15	91.71	58.57	77.81
Tailored	DAE	91.61	68.76	38.26	23.17	43.12	96.85	58.78	77.24
	SVDD	88.91	63.48	27.57	15.68	22.40	80.65	56.90	60.94
Self-supervision	PPRD	91.20	68.70	39.20	23.67	36.66	96.27	59.93	76.10
	SPRD	92.06	70.26	42.14	25.03	40.89	96.51	60.60	76.34
MDOD	MODDIS	66.47	45.73	15.15	7.98	11.48	22.98	52.75	56.74
	CAAE	90.77	68.73	25.38	14.68	23.06	63.65	57.15	68.35

TABLE VIII: AUPR-abnormal (%) of different baselines on image based multi-view datasets.

Type		MNIST	FashionMNIST	Cifar10	Cifar100	SVHN	Cat_vs_Dog	MedMNIST	MvTecAD
Fusion	SUM	99.66	99.02	96.86	98.00	95.99	98.97	85.42	95.75
	MAX	99.66	99.00	96.69	97.95	95.90	98.98	85.40	95.69
	NN	99.67	99.03	96.61	97.95	95.99	99.01	85.48	95.44
	TF	99.66	98.99	96.78	97.96	96.02	98.89	85.34	95.34
Alignment	DIS	99.66	99.01	96.80	97.98	96.09	98.96	85.56	95.40
	SIM	99.66	99.00	96.87	98.00	96.14	99.01	85.40	95.58
	DCCA	99.66	98.92	96.17	97.83	96.09	97.74	85.48	95.65
Tailored	DAE	99.66	99.00	96.87	97.98	96.17	99.01	85.37	95.45
	SVDD	99.61	98.92	95.91	97.60	94.90	91.63	84.57	91.82
Self-supervision	PPRD	99.65	99.02	96.71	97.91	95.63	98.91	85.68	95.65
	SPRD	99.67	99.04	96.86	97.95	95.85	99.00	85.70	95.56
MDOD	MODDIS	99.24	98.26	94.36	97.11	92.45	68.09	84.05	88.28
	CAAE	99.66	99.05	95.80	97.60	94.77	86.94	84.47	92.98

TABLE IX: TNR(%)@95%TPR of different baselines on image based multi-view datasets.

Type		MNIST	FashionMNIST	Cifar10	Cifar100	SVHN	Cat_vs_Dog	MedMNIST	MvTecAD
Fusion	SUM	81.72	66.76	24.88	22.55	19.53	90.43	43.77	66.27
	MAX	81.72	66.74	24.88	22.55	19.48	90.42	43.78	65.54
	NN	81.72	66.77	24.88	22.55	19.18	91.58	43.76	65.27
	TF	81.72	66.80	24.88	22.55	19.28	89.04	43.76	64.73
Alignment	DIS	81.72	66.74	24.88	22.55	19.12	92.13	43.58	64.10
	SIM	81.71	66.74	24.88	22.55	19.42	91.60	43.64	64.68
	DCCA	81.69	66.71	24.88	22.55	18.61	67.20	43.74	64.40
Tailored	DAE	81.72	66.74	24.88	22.55	19.64	91.95	43.64	64.11
	SVDD	82.37	66.87	23.17	20.45	19.86	30.84	39.62	56.36
Self-supervision	PPRD	81.72	66.79	24.88	22.55	18.81	89.15	43.80	65.65
	SPRD	81.69	66.81	24.88	22.55	19.09	90.41	43.94	65.76
MDOD	MODDIS	80.59	65.93	22.68	20.60	16.31	1.91	39.58	41.10
	CAAE	82.82	66.97	24.04	20.97	17.62	17.56	40.14	54.48

TABLE X: AUPR-normal (%) of different baselines on video based multi-view datasets.

Type		UCSDped1	UCSDped2	UMN_scene1	UMN_scene2	UMN_scene3	Avenue	ShanghaiTech
Fusion	SUM	97.95	96.23	99.39	96.69	97.42	98.13	91.72
	MAX	97.77	95.31	99.31	96.41	97.15	98.11	91.00
	NN	97.83	95.19	99.53	96.38	97.44	98.05	91.64
	TF	97.89	96.04	99.43	96.56	97.48	98.10	91.54
Alignment	DIS	97.66	94.72	99.11	96.28	97.30	98.25	90.51
	SIM	97.85	95.43	99.13	96.20	96.94	97.13	91.10
	DCCA	97.14	95.52	98.91	96.09	97.19	97.92	90.87
Tailored	DAE	97.58	94.59	99.03	96.48	96.72	98.11	90.85
	SVDD	94.68	96.14	99.67	92.16	98.62	98.15	86.38
Self-supervision	PPRD	97.08	97.30	99.64	95.92	98.96	97.76	84.73
	SPRD	97.17	96.85	99.66	96.94	98.57	98.11	87.18
MDOD	MODDIS	96.87	95.22	99.71	94.69	98.50	97.80	87.40
	CAAE	96.44	96.06	99.70	95.22	98.75	98.34	90.99

TABLE XI: AUPR-abnormal (%) of different baselines on video based multi-view datasets.

Type		UCSDped1	UCSDped2	UMN_scene1	UMN_scene2	UMN_scene3	Avenue	ShanghaiTech
Fusion	SUM	44.51	72.89	95.36	75.49	80.51	35.67	31.10
	MAX	37.86	68.74	94.61	73.89	79.33	31.29	27.41
	NN	39.99	68.09	95.03	74.97	80.68	33.58	30.77
	TF	44.50	72.34	95.37	75.37	81.09	36.02	30.47
Alignment	DIS	39.85	66.87	94.46	73.93	80.70	41.74	27.09
	SIM	44.31	69.54	94.27	73.85	79.04	33.23	28.44
	DCCA	33.84	66.77	93.03	73.01	78.91	35.26	28.98
Tailored	DAE	39.52	66.45	93.99	74.50	78.77	31.71	27.42
	SVDD	29.79	73.50	94.32	69.95	82.95	55.60	26.45
Self-supervision	PPRD	39.58	75.30	95.41	72.50	85.24	25.00	15.83
	SPRD	41.04	73.81	95.72	74.67	82.02	27.98	19.61
MDOD	MODDIS	37.13	65.66	94.82	73.00	82.97	37.05	19.39
	CAAE	31.94	68.48	95.44	72.00	82.97	49.24	41.31

TABLE XII: TNR(%)@95%TPR of different baselines on video based multi-view datasets.

Type		UCSDped1	UCSDped2	UMN_scene1	UMN_scene2	UMN_scene3	Avenue	ShanghaiTech
Fusion	SUM	40.12	61.68	93.87	67.04	77.01	54.68	17.01
	MAX	33.27	56.59	92.52	64.70	75.27	49.78	14.39
	NN	35.99	57.29	93.36	66.53	77.21	52.56	17.48
	TF	39.92	60.82	93.82	67.17	77.07	54.47	16.53
Alignment	DIS	35.37	55.24	92.69	64.94	78.13	46.01	13.82
	SIM	40.31	58.63	92.23	65.12	74.23	46.42	14.69
	DCCA	30.84	56.04	89.58	63.26	73.71	40.73	18.09
Tailored	DAE	35.27	54.10	91.81	66.09	75.30	48.79	14.15
	SVDD	23.68	63.64	93.91	60.25	81.25	59.90	16.30
Self-supervision	PPRD	38.47	64.21	94.16	63.01	83.13	35.54	4.13
	SPRD	39.90	63.33	94.29	65.70	79.75	38.75	8.52
MDOD	MODDIS	37.09	53.52	93.70	64.16	80.21	45.85	8.55
	CAAE	35.33	61.54	94.54	62.33	81.60	53.81	33.20

TABLE XIII: AUPR-normal (%) of different baselines on selected existing multi-view datasets with random train/test split.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
SUM	$75.85_{\pm 1.47}$	$42.86_{\pm 1.02}$	$98.98_{\pm 0.35}$	$36.16_{\pm 1.42}$	$43.67_{\pm 1.65}$	$13.87_{\pm 0.30}$	$39.71_{\pm 1.42}$	$1.34_{\pm 0.05}$	$2.69_{\pm 0.13}$	$17.47_{\pm 0.45}$
MAX	$75.85_{\pm 1.46}$	$43.44_{\pm 1.13}$	$98.89_{\pm 0.39}$	$36.19_{\pm 1.37}$	$43.59_{\pm 1.71}$	$13.80_{\pm 0.29}$	$39.76_{\pm 1.55}$	$1.40_{\pm 0.05}$	$2.82_{\pm 0.18}$	$17.20_{\pm 0.42}$
NN	$75.83_{\pm 1.46}$	$42.57_{\pm 1.11}$	$98.97_{\pm 0.38}$	$36.22_{\pm 1.37}$	$43.55_{\pm 1.69}$	$13.51_{\pm 0.37}$	$40.68_{\pm 1.45}$	$1.37_{\pm 0.07}$	$2.80_{\pm 0.19}$	$16.75_{\pm 0.55}$
TF	$75.82_{\pm 1.45}$	$42.26_{\pm 1.19}$	$92.31_{\pm 3.44}$	$36.43_{\pm 1.44}$	$43.74_{\pm 1.66}$	$13.99_{\pm 0.30}$	$41.80_{\pm 1.74}$	$1.33_{\pm 0.09}$	$2.50_{\pm 0.11}$	$16.29_{\pm 0.61}$
DIS	$75.84_{\pm 1.45}$	$43.01_{\pm 1.10}$	$99.09_{\pm 0.29}$	$36.26_{\pm 1.39}$	$43.67_{\pm 1.71}$	$13.99_{\pm 0.31}$	$33.76_{\pm 1.96}$	$1.35_{\pm 0.04}$	$2.51_{\pm 0.16}$	$16.20_{\pm 0.39}$
SIM	$75.84_{\pm 1.44}$	$42.46_{\pm 1.06}$	$98.93_{\pm 0.32}$	$36.42_{\pm 1.34}$	$43.81_{\pm 1.74}$	$14.15_{\pm 0.31}$	$32.24_{\pm 2.16}$	$1.31_{\pm 0.06}$	$2.44_{\pm 0.13}$	$16.17_{\pm 0.67}$
DCCA	$75.86_{\pm 1.46}$	$42.77_{\pm 1.09}$	$98.74_{\pm 0.44}$	$36.38_{\pm 1.29}$	$43.80_{\pm 1.70}$	$14.15_{\pm 0.34}$	$34.26_{\pm 1.48}$	$1.32_{\pm 0.05}$	$2.52_{\pm 0.22}$	$16.00_{\pm 0.49}$
DAE	$75.84_{\pm 1.45}$	$42.89_{\pm 1.14}$	$99.05_{\pm 0.33}$	$36.31_{\pm 1.44}$	$43.93_{\pm 1.62}$	$13.84_{\pm 0.28}$	$31.63_{\pm 1.93}$	$1.33_{\pm 0.04}$	$2.38_{\pm 0.10}$	$16.21_{\pm 0.44}$
SVDD	$73.41_{\pm 1.37}$	$28.02_{\pm 2.22}$	$90.94_{\pm 1.04}$	$27.23_{\pm 1.45}$	$40.41_{\pm 2.01}$	$16.94_{\pm 0.82}$	$31.23_{\pm 1.58}$	$1.22_{\pm 0.04}$	$2.34_{\pm 0.15}$	$5.65_{\pm 0.39}$
PPRD	$75.84_{\pm 1.45}$	$45.04_{\pm 1.11}$	$98.41_{\pm 0.51}$	$36.13_{\pm 1.41}$	$43.55_{\pm 1.71}$	$13.48_{\pm 0.31}$	$45.11_{\pm 1.35}$	$1.37_{\pm 0.05}$	$2.63_{\pm 0.14}$	$16.96_{\pm 0.56}$
SPRD	$75.83_{\pm 1.47}$	$42.30_{\pm 1.07}$	$98.66_{\pm 0.53}$	$36.24_{\pm 1.36}$	$43.52_{\pm 1.66}$	$13.40_{\pm 0.32}$	$46.33_{\pm 1.18}$	$1.39_{\pm 0.04}$	$2.98_{\pm 0.24}$	$18.13_{\pm 0.36}$
MODDIS	$73.78_{\pm 1.24}$	$16.84_{\pm 1.34}$	$44.56_{\pm 1.71}$	$30.79_{\pm 1.21}$	$41.65_{\pm 1.65}$	$14.15_{\pm 0.35}$	$38.72_{\pm 1.69}$	$0.99_{\pm 0.02}$	$2.08_{\pm 0.07}$	$1.97_{\pm 0.07}$
CAAE	$71.82_{\pm 1.44}$	$26.99_{\pm 1.39}$	$96.21_{\pm 0.67}$	$28.33_{\pm 1.35}$	$40.50_{\pm 1.63}$	$15.06_{\pm 0.46}$	$35.19_{\pm 1.51}$	$1.26_{\pm 0.04}$	$2.54_{\pm 0.13}$	$7.15_{\pm 0.56}$

TABLE XIV: AUPR-abnormal (%) of different baselines on selected existing multi-view datasets with random train/test split.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
SUM	$99.40_{\pm 0.07}$	$97.80_{\pm 0.14}$	$99.97_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.41_{\pm 0.06}$	$99.55_{\pm 0.01}$	$99.09_{\pm 0.01}$	$99.53_{\pm 0.02}$
MAX	$99.40_{\pm 0.07}$	$97.80_{\pm 0.16}$	$99.95_{\pm 0.04}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.42_{\pm 0.05}$	$99.55_{\pm 0.01}$	$99.11_{\pm 0.02}$	$99.52_{\pm 0.02}$
NN	$99.40_{\pm 0.07}$	$97.77_{\pm 0.14}$	$99.97_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.41_{\pm 0.03}$	$99.55_{\pm 0.01}$	$99.11_{\pm 0.02}$	$99.52_{\pm 0.01}$
TF	$99.40_{\pm 0.07}$	$97.76_{\pm 0.15}$	$99.62_{\pm 0.24}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.43_{\pm 0.07}$	$99.55_{\pm 0.01}$	$99.08_{\pm 0.01}$	$99.51_{\pm 0.02}$
DIS	$99.40_{\pm 0.07}$	$97.89_{\pm 0.16}$	$99.98_{\pm 0.01}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.27_{\pm 0.06}$	$99.55_{\pm 0.01}$	$99.05_{\pm 0.02}$	$99.51_{\pm 0.02}$
SIM	$99.40_{\pm 0.07}$	$97.88_{\pm 0.15}$	$99.97_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.51_{\pm 0.07}$	$99.25_{\pm 0.06}$	$99.54_{\pm 0.01}$	$99.05_{\pm 0.02}$	$99.50_{\pm 0.02}$
DCCA	$99.40_{\pm 0.07}$	$97.88_{\pm 0.15}$	$99.97_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.51_{\pm 0.07}$	$99.33_{\pm 0.06}$	$99.55_{\pm 0.01}$	$99.05_{\pm 0.02}$	$99.50_{\pm 0.02}$
DAE	$99.40_{\pm 0.07}$	$97.90_{\pm 0.15}$	$99.98_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.23_{\pm 0.05}$	$99.55_{\pm 0.01}$	$99.03_{\pm 0.02}$	$99.51_{\pm 0.01}$
SVDD	$99.33_{\pm 0.07}$	$97.29_{\pm 0.23}$	$99.74_{\pm 0.05}$	$97.04_{\pm 0.11}$	$98.10_{\pm 0.13}$	$96.39_{\pm 0.09}$	$99.14_{\pm 0.05}$	$99.53_{\pm 0.01}$	$99.17_{\pm 0.05}$	$98.81_{\pm 0.08}$
PPRD	$99.40_{\pm 0.07}$	$97.69_{\pm 0.17}$	$99.92_{\pm 0.05}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.55_{\pm 0.04}$	$99.55_{\pm 0.01}$	$99.12_{\pm 0.01}$	$99.48_{\pm 0.01}$
SPRD	$99.40_{\pm 0.07}$	$97.49_{\pm 0.17}$	$99.93_{\pm 0.04}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.53_{\pm 0.04}$	$99.56_{\pm 0.01}$	$99.14_{\pm 0.02}$	$99.53_{\pm 0.02}$
MODDIS	$99.35_{\pm 0.07}$	$94.57_{\pm 0.21}$	$97.33_{\pm 0.21}$	$97.79_{\pm 0.06}$	$98.68_{\pm 0.10}$	$96.43_{\pm 0.07}$	$99.22_{\pm 0.10}$	$99.51_{\pm 0.01}$	$98.90_{\pm 0.03}$	$97.62_{\pm 0.07}$
CAAE	$99.24_{\pm 0.05}$	$97.24_{\pm 0.18}$	$99.90_{\pm 0.03}$	$97.34_{\pm 0.09}$	$98.46_{\pm 0.11}$	$96.41_{\pm 0.09}$	$99.33_{\pm 0.05}$	$99.55_{\pm 0.01}$	$99.19_{\pm 0.04}$	$99.06_{\pm 0.03}$

TABLE XV: TNR(%)@95%TPR of different baselines on selected existing multi-view datasets with random train/test split.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
SUM	$67.82_{\pm 3.12}$	$33.88_{\pm 2.94}$	$99.77_{\pm 0.20}$	$34.13_{\pm 1.80}$	$49.94_{\pm 3.05}$	$24.20_{\pm 0.72}$	$62.45_{\pm 2.44}$	$12.91_{\pm 0.89}$	$21.93_{\pm 1.45}$	$44.33_{\pm 1.47}$
MAX	$67.82_{\pm 3.12}$	$33.64_{\pm 2.81}$	$99.72_{\pm 0.20}$	$34.14_{\pm 1.82}$	$50.08_{\pm 3.08}$	$24.20_{\pm 0.72}$	$62.86_{\pm 2.23}$	$13.01_{\pm 1.02}$	$21.78_{\pm 1.41}$	$43.26_{\pm 2.09}$
NN	$67.81_{\pm 3.13}$	$33.01_{\pm 2.60}$	$99.75_{\pm 0.31}$	$34.18_{\pm 1.84}$	$50.00_{\pm 3.12}$	$24.20_{\pm 0.72}$	$62.73_{\pm 2.34}$	$12.96_{\pm 0.88}$	$22.02_{\pm 1.62}$	$43.68_{\pm 1.35}$
TF	$67.81_{\pm 3.12}$	$32.81_{\pm 3.24}$	$89.34_{\pm 7.64}$	$34.16_{\pm 1.84}$	$50.00_{\pm 3.06}$	$24.20_{\pm 0.72}$	$62.82_{\pm 3.09}$	$12.99_{\pm 0.81}$	$21.17_{\pm 1.52}$	$43.35_{\pm 2.19}$
DIS	$67.82_{\pm 3.12}$	$35.09_{\pm 2.95}$	$99.78_{\pm 0.12}$	$34.13_{\pm 1.81}$	$50.02_{\pm 3.01}$	$24.20_{\pm 0.72}$	$56.84_{\pm 3.25}$	$12.81_{\pm 0.82}$	$21.04_{\pm 1.48}$	$42.03_{\pm 1.44}$
SIM	$67.82_{\pm 3.13}$	$35.41_{\pm 3.78}$	$99.79_{\pm 0.12}$	$34.13_{\pm 1.81}$	$49.95_{\pm 3.04}$	$24.20_{\pm 0.72}$	$56.08_{\pm 2.90}$	$12.82_{\pm 0.75}$	$21.18_{\pm 1.31}$	$42.30_{\pm 3.50}$
DCCA	$67.82_{\pm 3.13}$	$35.85_{\pm 3.83}$	$99.63_{\pm 0.34}$	$34.10_{\pm 1.82}$	$50.02_{\pm 3.12}$	$24.20_{\pm 0.72}$	$59.05_{\pm 3.18}$	$12.89_{\pm 0.83}$	$21.09_{\pm 1.70}$	$42.20_{\pm 2.43}$
DAE	$67.82_{\pm 3.12}$	$35.65_{\pm 3.14}$	$99.83_{\pm 0.09}$	$34.13_{\pm 1.81}$	$49.98_{\pm 3.09}$	$24.20_{\pm 0.72}$	$56.40_{\pm 2.90}$	$12.76_{\pm 0.78}$	$20.93_{\pm 1.67}$	$42.10_{\pm 1.20}$
SVDD	$66.32_{\pm 4.10}$	$28.64_{\pm 3.25}$	$88.91_{\pm 1.95}$	$18.64_{\pm 2.33}$	$32.48_{\pm 3.35}$	$20.68_{\pm 1.34}$	$52.17_{\pm 2.56}$	$13.05_{\pm 1.20}$	$18.95_{\pm 1.18}$	$16.82_{\pm 1.17}$
PPRD	$67.82_{\pm 3.12}$	$29.66_{\pm 3.54}$	$99.44_{\pm 0.36}$	$34.13_{\pm 1.80}$	$50.02_{\pm 3.07}$	$24.20_{\pm 0.72}$	$70.22_{\pm 2.62}$	$12.94_{\pm 0.84}$	$21.68_{\pm 1.39}$	$38.97_{\pm 1.99}$
SPRD	$67.82_{\pm 3.13}$	$27.81_{\pm 2.42}$	$99.66_{\pm 0.25}$	$34.08_{\pm 1.80}$	$50.07_{\pm 3.06}$	$24.20_{\pm 0.72}$	$69.30_{\pm 2.43}$	$13.17_{\pm 0.87}$	$22.31_{\pm 1.88}$	$43.60_{\pm 2.74}$
MODDIS	$66.26_{\pm 3.08}$	$13.18_{\pm 1.29}$	$33.24_{\pm 3.99}$	$25.72_{\pm 1.52}$	$45.48_{\pm 3.24}$	$22.40_{\pm 0.82}$	$56.43_{\pm 4.52}$	$13.13_{\pm 1.19}$	$19.54_{\pm 1.13}$	$6.68_{\pm 0.44}$
CAAE	$63.68_{\pm 3.38}$	$26.91_{\pm 2.20}$	$97.44_{\pm 0.89}$	$21.30_{\pm 1.64}$	$40.56_{\pm 4.08}$	$21.05_{\pm 1.43}$	$60.12_{\pm 2.34}$	$13.72_{\pm 1.40}$	$20.11_{\pm 1.50}$	$21.69_{\pm 1.65}$

TABLE XVI: AUPR-normal (%) of different late fusion strategies on selected existing multi-view datasets.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
LF-AVG	$75.84_{\pm 1.45}$	$42.89_{\pm 1.14}$	$99.05_{\pm 0.33}$	$36.31_{\pm 1.44}$	$43.93_{\pm 1.62}$	$13.84_{\pm 0.28}$	$31.63_{\pm 1.93}$	$1.33_{\pm 0.04}$	$2.38_{\pm 0.10}$	$16.21_{\pm 0.44}$
LF-MIN	$72.62_{\pm 1.59}$	$43.20_{\pm 1.09}$	$97.05_{\pm 0.58}$	$31.54_{\pm 1.17}$	$34.60_{\pm 1.58}$	$13.59_{\pm 0.29}$	$18.22_{\pm 1.24}$	$1.24_{\pm 0.06}$	$2.08_{\pm 0.09}$	$14.23_{\pm 0.57}$
LF-MAX	$65.51_{\pm 0.93}$	$27.30_{\pm 7.88}$	$66.77_{\pm 3.09}$	$19.94_{\pm 3.08}$	$22.33_{\pm 3.65}$	$10.31_{\pm 0.15}$	$46.49_{\pm 1.95}$	$0.85_{\pm 0.02}$	$1.98_{\pm 0.06}$	$11.08_{\pm 0.50}$

TABLE XVII: AUPR-abnormal (%) of different late fusion strategies on selected existing multi-view datasets.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
LF-AVG	$99.40_{\pm 0.07}$	$97.90_{\pm 0.15}$	$99.98_{\pm 0.02}$	$98.36_{\pm 0.06}$	$98.74_{\pm 0.09}$	$96.50_{\pm 0.07}$	$99.23_{\pm 0.05}$	$99.55_{\pm 0.01}$	$99.03_{\pm 0.02}$	$99.51_{\pm 0.01}$
LF-MIN	$99.27_{\pm 0.08}$	$97.90_{\pm 0.18}$	$99.95_{\pm 0.03}$	$98.19_{\pm 0.07}$	$98.20_{\pm 0.11}$	$96.24_{\pm 0.07}$	$98.88_{\pm 0.08}$	$99.52_{\pm 0.01}$	$98.92_{\pm 0.02}$	$99.40_{\pm 0.02}$
LF-MAX	$99.42_{\pm 0.04}$	$93.61_{\pm 0.20}$	$98.80_{\pm 0.18}$	$95.08_{\pm 0.10}$	$95.73_{\pm 0.08}$	$95.99_{\pm 0.04}$	$99.50_{\pm 0.06}$	$99.44_{\pm 0.01}$	$99.11_{\pm 0.02}$	$99.45_{\pm 0.01}$

TABLE XVIII: TNR(%)@95%TPR of different late fusion strategies on selected existing multi-view datasets.

	BBC	BDGP	Caltech20	Citeseer	Cora	Reuters	Wiki	AwA	NUS-Wide	SunRGBD
LF-AVG	$67.82_{\pm 3.12}$	$35.65_{\pm 3.14}$	$99.83_{\pm 0.09}$	$34.13_{\pm 1.81}$	$49.98_{\pm 3.09}$	$24.20_{\pm 0.72}$	$56.40_{\pm 2.90}$	$12.76_{\pm 0.78}$	$20.93_{\pm 1.67}$	$42.10_{\pm 1.20}$
LF-MIN	$62.41_{\pm 3.59}$	$34.18_{\pm 3.67}$	$98.03_{\pm 1.78}$	$35.46_{\pm 2.34}$	$36.23_{\pm 2.32}$	$20.12_{\pm 0.77}$	$48.37_{\pm 2.42}$	$11.51_{\pm 0.54}$	$17.98_{\pm 1.28}$	$35.70_{\pm 3.22}$
LF-MAX	$71.53_{\pm 3.39}$	$9.31_{\pm 0.64}$	$61.33_{\pm 3.18}$	$23.42_{\pm 0.70}$	$31.19_{\pm 1.07}$	$18.63_{\pm 0.77}$	$69.33_{\pm 3.38}$	$12.30_{\pm 0.67}$	$17.27_{\pm 1.48}$	$41.28_{\pm 0.94}$

TABLE XIX: AUPR-normal (%) of different baselines on existing multi-view/multi-modal datasets with given train/test set split.

Type		YtFace	CMU-MOSEI	DirverAD
Fusion	SUM	50.54	18.08	97.59
	MAX	50.02	18.05	97.31
	NN	50.17	18.06	95.94
	TF	41.18	18.00	97.52
Alignment	DIS	50.27	16.25	97.74
	SIM	50.59	18.04	97.81
	DCCA	43.00	17.83	97.77
Tailored	DAE	49.91	18.14	97.08
	DSV	50.72	16.81	88.55
Self-supervision	PPRD	50.82	17.81	95.23
	SPRD	49.66	17.82	96.69
MDOD	MODDIS	30.69	17.75	98.14
	CAAE	49.86	20.95	95.59

TABLE XX: AUPR-abnormal (%) of different baselines on existing multi-view/multi-modal datasets with given train/test set split.

Type		YtFace	CMU-MOSEI	DirverAD
Fusion	SUM	99.48	83.69	93.46
	MAX	99.47	83.64	91.37
	NN	99.48	83.73	81.61
	TF	99.42	83.74	93.57
Alignment	DIS	99.49	83.44	94.05
	SIM	99.49	83.77	94.45
	DCCA	99.47	83.86	94.66
Tailored	DAE	99.49	83.92	89.52
	DSV	99.56	83.26	77.20
Self-supervision	PPRD	99.49	83.79	83.33
	SPRD	99.46	83.67	88.92
MDOD	MODDIS	99.29	83.77	95.50
	CAAE	99.52	86.36	81.15

TABLE XXI: TNR(%)@95%TPR of different baselines on existing multi-view/multi-modal datasets with given train/test set split.

Type		YtFace	CMU-MOSEI	DirverAD
Fusion	SUM	43.54	12.30	82.65
	MAX	42.00	12.27	78.37
	NN	43.34	13.28	43.63
	TF	41.31	12.15	81.09
Alignment	DIS	43.53	12.81	84.51
	SIM	43.70	13.03	85.43
	DCCA	43.41	14.02	86.56
Tailored	DAE	43.34	13.99	77.64
	DSV	49.03	8.18	51.03
Self-supervision	PPRD	42.83	13.56	56.94
	SPRD	42.81	13.15	68.29
MDOD	MODDIS	38.05	12.90	88.58
	CAAE	44.03	15.39	52.51