APPENDIX

.1 Details of Dataset Description

We show the detailed description of 25 publicly available real-world datasets in Table A1, which include several domains such as disease diagnosis, speech recognition, and image identification.

Table A1: Dataset description.

Dataset	N	D	#anomalies	#anomaly ratio (%)
ALOI	49534	27	1508	3.04
annthyroid	7200	6	534	7.42
Cardiotocography	2114	21	466	22.04
fault	1941	27	673	34.67
http	567498	3	2211	0.39
landsat	6435	36	1333	20.71
letter	1600	32	100	6.25
magic.gamma	19020	10	6688	35.16
mammography	11183	6	260	2.32
mnist	7603	100	700	9.21
musk	3062	166	97	3.17
optdigits	5216	64	150	2.88
PageBlocks	5393	10	510	9.46
pendigits	6870	16	156	2.27
satellite	6435	36	2036	31.64
satimage-2	5803	36	71	1.22
shuttle	49097	9	3511	7.15
skin	245057	3	50859	20.75
SpamBase	4207	57	1679	39.91
speech	3686	400	61	1.65
thyroid	3772	6	93	2.47
vowels	1456	12	50	3.43
Waveform	3443	21	100	2.90
Wilt	4819	5	257	5.33
yeast	1484	8	507	34.16

.2 Additional Training Details

For unsupervised baselines, Iforest, ECOD, and DeepSVDD are built using the PyOD [71] library. Labeled anomalies are combined with unlabeled data for constructing the validation set, in order to tune the hyperparameters of these unsupervised methods via the grid search method, since tuning their hyperparameters on a small validation set often yields better performance than using the default settings [57]. Table A2 shows the hyperparameter grids, where ECOD is not considered since it is a parameter-free method.

Table A2: Hyperparameter grid of the unsupervised models.

Model	Hyperparameter
Iforest	n_estimators: [10, 50, 100, 500]
DeepSVDD	epochs: [20, 50, 100, 200]

We replace the convolutional layer in the original GANomaly with the dense layer for evaluating it on the tabular data, where the hidden size of the encoder-decoder-encoder structure of GANomaly is set to half of the input dimension. We realize the PReNet in PyTorch as we do not find the open-source codes, and set the hyperparameters in PReNet according to its original paper. Other models are built based on their corresponding source codes. If not specified, we train these baseline models according to their default hyperparameters mentioned in the original papers.

.3 Experimental Results of Selecting Intersection Points

We demonstrate the experimental results of different intersection point selection strategies in Table A3. The detection performance of a randomly selected intersection point is very close to that of ensembling all the calculated intersection points w.r.t. different ratios of labeled anomalies $\gamma_l = 5\%$, $\gamma_l = 10\%$ and $\gamma_l = 20\%$. This is due to the fact that the calculation of the overlap area repeats many times (epochs×batchsize), which is essentially similar to the average results of the ensemble strategy. Moreover, random sampling of the intersection points can improve computational efficiency since the overlap area only needs to be estimated once in a training batch.

Table A3: AUC-ROC and AUC-PR results of different intersection point selection strategies. MLP-Overlap corresponds to the default strategy that randomly chooses one of the intersection points for estimating the overlap area via Eq.9. MLP-Overlap-E refers to the ensemble strategy by taking the average of the overlap areas calculated based on each intersection point.

(a) AUC-ROC results.

	$\gamma_l = 5\%$	$\gamma_l = 10\%$	$\gamma_l = 20\%$
MLP-Overlap	0.847±0.145	0.880±0.132	0.893±0.133
MLP-Overlap-E	0.849±0.146	0.879±0.132	0.894±0.131

(b) AUC-PR results.

	$\gamma_l = 5\%$	$\gamma_l = 10\%$	$\gamma_l = 20\%$
MLP-Overlap	0.623±0.291	0.674±0.286	0.696±0.288
MLP-Overlap-E	0.626±0.291	0.673±0.285	0.694±0.287

.4 Detailed Experimental Results

We show the AUC-ROC results of model performance in Table A4 and ablation study in Table A5, corresponding to Section 4.2.1 and 4.2.3 in the main paper, respectively. These experimental results are basically consistent with the main paper. Besides, we show the detailed results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 5\%$, $\gamma_l = 10\%$ and $\gamma_l = 20\%$ in Table A6~A11. The best performing method(s) is marked in **bold**.

Table A4: Average AUC-ROC performance over 25 real-world datasets. Each experiment is repeated 5 times. γ_l stands for the ratio of labeled anomalies to all true anomalies in the training set. Δ Perf. shows the relative improvement of Overlap loss based models over their corresponding counterparts. ***, ** and * denote statistical significance at 1%, 5% and 10% of Wilcoxon signed rank test, respectively. The best results are in bold.

Architecture	Model	Supervision	$\gamma_l = 5\%$		$\gamma_l = 10\%$		$\gamma_l = 20\%$	
			AUC-ROC	Δ Perf.	AUC-ROC	Δ Perf.	AUC-ROC	Δ Perf.
Typical	Iforest	Unsup	0.737±0.187	/	0.737±0.187	/	0.737±0.187	/
	ECOD	Unsup	0.701±0.208	/	0.701±0.208	/	0.701±0.208	/
	DeepSVDD	Unsup	0.504±0.028	/	0.504±0.028	/	0.504±0.028	/
	GANomaly	Semi	0.655±0.162	/	0.648±0.153	/	0.665±0.152	/
	DeepSAD	Semi	0.823±0.142	/	0.859±0.136	/	0.888±0.129	/
	REPEN	Weak	0.810±0.166	/	0.832±0.165	/	0.848±0.163	/
MLP	DevNet	Weak	0.842±0.148	+0.57%	0.861±0.135	+2.16%	0.873±0.129	+2.40%
	PReNet	Weak	0.846±0.146	+0.18%	0.866±0.132	+1.61%	0.876±0.127	+1.97%
	MLP-Overlap (ours)	Weak	**0.847±0.145**	/	**0.880±0.132**	/	**0.893±0.133**	/
AutoEncoder	FEAWAD	Sup	0.771±0.211	+11.72%***	0.849±0.133	+4.34%***	0.876±0.133	+2.42%***
	FEAWAD	Weak	0.808±0.154	+6.60%***	0.848±0.145	+4.51%***	0.876±0.129	+2.43%***
	AE-Overlap (ours)	Weak	**0.862±0.144**	/	**0.886±0.137**	/	**0.897±0.132**	/
ResNet	ResNet	Sup	0.651±0.158	+28.48%***	0.736±0.124	+19.74%***	0.816±0.127	+11.03%***
	ResNet-Overlap (ours)	Weak	**0.836±0.146**	/	**0.882±0.134**	/	**0.906±0.122**	/
Transformer	FTTransformer	Sup	0.827±0.159	+3.00%**	0.859±0.146	+1.80%	0.889±0.129	+1.23%
	FTTransformer-Overlap (ours)	Weak	**0.851±0.138**	/	**0.874±0.130**	/	**0.900±0.127**	/

Table A5: AUC-ROC results of ablation studies. Overlap-Gaussian refers to the basic method mentioned in Section 3.3.1. Overlap-Arbitrary refers to the basic method of Eq.4. Overlap-Ranking isolates the ranking loss in Eq.5. Overlap-Combined corresponds to the combined loss form of both Overlap-Arbitrary and Overlap-Ranking as illustrated in Eq.5. Overlap-Proposed refers to the final solution in this paper.

Method	$\gamma_l = 5\%$					$\gamma_l = 10\%$					$\gamma_l = 20\%$				
	VAE	**MLP**	**AE**	**ResNet**	**FTT**	**VAE**	**MLP**	**AE**	**ResNet**	**FTT**	**VAE**	**MLP**	**AE**	**ResNet**	**FTT**
Overlap-Gaussian	0.539	/	/	/	/	0.540	/	/	/	/	0.541	/	/	/	/
Overlap-Arbitrary	/	0.496	0.531	0.493	0.521	/	0.498	0.543	0.534	0.482	/	0.502	0.516	0.541	0.483
Overlap-Ranking	/	0.810	0.822	0.807	**0.862**	/	0.845	0.857	0.855	**0.888**	/	0.873	0.874	0.890	**0.906**
Overlap-Combined	/	0.843	0.854	0.820	0.610	/	**0.881**	0.885	0.874	0.602	/	**0.898**	**0.901**	**0.908**	0.589
Overlap-Proposed	/	**0.847**	**0.862**	**0.836**	0.851	/	0.880	**0.886**	**0.882**	0.874	/	0.893	0.897	0.906	0.900

Table A6: AUC-ROC results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 5\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.548	0.560	0.525	0.552	**0.576**	0.528	0.486	0.492	0.506	0.497	0.547	0.532	0.483	0.500	0.501	0.515
annthyroid	0.828	0.783	0.502	0.769	0.818	0.803	0.805	0.822	0.882	0.798	0.903	0.914	0.761	0.872	0.922	**0.985**
Cardiotocography	0.692	0.684	0.514	0.575	0.807	0.898	**0.911**	0.908	0.874	0.818	0.791	0.883	0.627	0.884	0.842	0.835
fault	0.569	0.444	0.485	0.628	**0.704**	0.694	0.696	0.683	0.645	0.671	0.657	0.668	0.502	0.665	0.657	0.656
http	0.999	0.981	0.509	0.980	0.998	0.999	0.999	0.999	**1.000**	0.000	0.999	0.999	0.835	**1.000**	0.999	**1.000**
landsat	0.483	0.571	0.501	0.514	0.854	0.584	0.776	0.779	0.833	0.767	0.802	0.816	0.671	0.735	**0.864**	**0.864**
letter	0.635	0.526	0.468	0.673	0.703	0.629	0.590	0.582	0.591	0.557	0.562	0.566	**0.752**	0.634	0.513	0.560
magic.gamma	0.732	0.648	0.498	0.580	0.819	0.799	0.827	0.832	0.840	0.646	0.818	0.842	0.695	0.824	0.812	**0.843**
mammography	0.861	0.909	0.520	0.781	0.915	0.919	0.924	0.919	0.912	0.853	0.919	**0.931**	0.748	0.867	0.900	0.921
mnist	0.803	0.846	0.523	0.705	0.862	0.917	**0.949**	0.925	0.823	0.935	0.869	0.925	0.586	0.720	0.916	0.870
musk	0.999	0.952	0.550	0.781	0.917	0.915	**1.000**	**1.000**	**1.000**	0.998	0.927	**1.000**	0.426	**1.000**	0.999	0.989
optdigits	0.674	0.612	0.522	0.384	0.934	0.986	**1.000**	0.999	0.995	0.966	0.988	0.999	0.654	0.994	0.977	0.916
PageBlocks	0.894	0.913	0.522	0.654	**0.934**	0.912	0.864	0.883	0.896	0.785	0.895	0.890	0.714	0.910	0.842	0.888
pendigits	0.955	0.910	0.485	0.707	0.965	0.997	0.996	0.993	0.995	0.958	0.997	**0.999**	0.682	0.994	0.989	0.984
satellite	0.699	0.750	0.503	0.722	0.883	0.807	0.853	0.852	0.852	0.766	0.834	**0.910**	0.740	0.907	0.874	0.880
satimage-2	**0.992**	0.966	0.525	0.969	0.981	0.986	0.991	0.989	0.970	0.921	0.967	0.988	0.367	0.973	0.942	0.932
shuttle	**0.996**	0.995	0.508	0.744	0.990	0.989	0.979	0.978	0.981	0.976	0.980	0.982	0.973	0.979	0.976	0.977
skin	0.684	0.391	0.500	0.542	0.995	0.919	0.951	0.954	0.982	**0.999**	0.978	0.987	0.998	0.994	0.993	0.965
SpamBase	0.633	0.660	0.500	0.534	0.690	0.838	0.902	0.909	0.876	0.748	0.768	0.841	0.598	0.769	0.870	**0.917**
speech	0.498	0.510	0.549	0.481	0.531	0.582	0.604	0.631	0.589	0.587	0.475	0.624	0.532	0.581	0.551	**0.648**
thyroid	0.981	0.979	0.521	0.919	0.941	0.990	0.994	**0.995**	0.981	0.863	0.818	0.988	0.387	0.954	0.980	0.968
vowels	0.765	0.440	0.407	0.792	0.767	0.812	0.847	0.891	0.860	0.777	0.665	**0.916**	0.646	0.787	0.735	0.823
Waveform	0.693	0.723	0.495	0.545	0.691	0.763	0.806	0.809	0.764	**0.882**	0.626	0.780	0.563	0.765	0.772	0.814
Wilt	0.427	0.395	0.502	0.388	0.799	0.529	0.689	0.695	0.922	0.894	0.819	0.945	0.795	**0.950**	0.658	0.910
yeast	0.382	0.387	0.477	0.449	0.512	0.467	0.625	0.627	0.611	0.625	0.610	0.619	0.530	**0.642**	0.583	0.627

Table A7: AUC-ROC results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 10\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.548	0.560	0.525	0.556	**0.574**	0.546	0.510	0.514	0.523	0.487	0.531	0.519	0.476	0.496	0.506	0.524
annthyroid	0.828	0.783	0.502	0.733	0.884	0.824	0.826	0.834	0.939	0.880	0.897	0.968	0.905	0.932	**0.990**	**0.990**
Cardiotocography	0.692	0.684	0.514	0.578	0.868	0.916	**0.931**	**0.931**	0.927	0.849	0.835	0.920	0.689	0.928	0.893	0.885
fault	0.569	0.444	0.485	0.631	**0.728**	0.720	0.724	0.719	0.695	0.692	0.674	0.695	0.570	0.721	0.694	0.699
http	0.999	0.981	0.509	0.785	0.999	**1.000**	0.999	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.829	**1.000**	**1.000**	**1.000**
landsat	0.483	0.571	0.501	0.522	**0.897**	0.561	0.779	0.789	0.878	0.805	0.805	0.839	0.743	0.836	0.891	0.891
letter	0.635	0.526	0.468	0.673	0.723	**0.753**	0.699	0.713	0.694	0.699	0.613	0.656	0.749	0.720	0.612	0.636
magic.gamma	0.732	0.648	0.498	0.611	0.847	0.809	0.827	0.833	0.870	0.688	0.841	0.874	0.755	0.871	0.834	**0.879**
mammography	0.861	0.909	0.520	0.774	0.907	0.925	0.926	0.924	0.931	0.793	0.918	**0.935**	0.812	0.932	0.908	0.919
mnist	0.803	0.846	0.523	0.707	0.916	0.966	**0.974**	0.963	0.915	0.957	0.940	0.962	0.753	0.911	0.947	0.939
musk	0.999	0.952	0.550	0.879	0.968	0.971	**1.000**	**1.000**	**1.000**	0.997	**1.000**	**1.000**	0.654	**1.000**	0.996	0.995
optdigits	0.674	0.612	0.522	0.383	0.979	0.993	**1.000**	0.998	0.996	0.928	0.982	0.999	0.652	**1.000**	0.985	0.903
PageBlocks	0.894	0.913	0.522	0.681	**0.945**	0.925	0.863	0.880	0.919	0.829	0.939	0.927	0.777	0.919	0.894	0.897
pendigits	0.955	0.910	0.485	0.713	0.993	0.996	0.995	0.995	0.995	0.981	0.992	**0.999**	0.761	**0.999**	0.997	0.981
satellite	0.699	0.750	0.503	0.726	0.909	0.807	0.851	0.849	0.899	0.840	0.852	**0.919**	0.759	0.910	0.908	0.891
satimage-2	**0.992**	0.966	0.525	0.969	0.986	0.989	0.988	0.986	0.954	0.959	0.971	0.978	0.801	0.987	0.974	0.949
shuttle	**0.996**	0.995	0.508	0.650	0.992	0.987	0.979	0.980	0.982	0.983	0.985	0.982	0.981	0.979	0.981	0.979
skin	0.684	0.391	0.500	0.516	0.997	0.910	0.951	0.954	0.992	**0.999**	0.992	0.996	0.998	0.993	0.995	0.985
SpamBase	0.633	0.660	0.500	0.538	0.801	0.861	0.915	0.928	0.921	0.812	0.862	0.916	0.677	0.893	0.914	**0.951**
speech	0.498	0.510	0.549	0.481	0.538	0.541	0.659	0.689	0.616	0.668	0.512	0.647	0.548	0.614	0.610	**0.698**
thyroid	0.981	0.979	0.521	0.920	0.965	0.994	**0.996**	**0.996**	0.994	0.974	0.897	0.988	0.698	**0.996**	0.993	0.995
vowels	0.765	0.440	0.407	0.794	0.805	0.896	0.926	0.942	0.925	0.906	0.899	**0.964**	0.674	0.952	0.753	0.917
Waveform	0.693	0.723	0.495	0.545	0.765	0.846	0.881	0.884	0.836	**0.887**	0.728	0.854	0.666	0.786	0.863	0.774
Wilt	0.427	0.395	0.502	0.391	0.918	0.597	0.691	0.698	0.944	0.950	0.876	0.969	0.890	**0.988**	0.685	0.942
yeast	0.382	0.387	0.477	0.448	0.576	0.460	0.638	0.644	0.648	0.663	0.649	0.639	0.596	**0.682**	0.655	0.642

Table A8: AUC-ROC results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 20\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.548	0.560	0.525	0.550	0.591	0.532	0.522	0.520	0.525	0.504	**0.596**	0.544	0.495	0.535	0.526	0.515
annthyroid	0.828	0.783	0.502	0.781	0.930	0.825	0.825	0.835	0.951	0.955	0.941	0.970	0.957	0.958	**0.992**	0.989
Cardiotocography	0.692	0.684	0.514	0.583	0.913	0.930	0.946	0.945	0.942	0.896	0.894	0.946	0.778	**0.948**	0.930	0.933
fault	0.569	0.444	0.485	0.633	**0.749**	0.743	0.733	0.738	0.698	0.669	0.718	0.673	0.642	0.745	0.713	0.717
http	0.999	0.981	0.509	0.783	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
landsat	0.483	0.571	0.501	0.538	**0.921**	0.560	0.789	0.782	0.899	0.872	0.846	0.806	0.806	0.884	0.913	0.901
letter	0.635	0.526	0.468	0.675	0.751	**0.806**	0.748	0.767	0.739	0.717	0.677	0.683	0.739	0.779	0.705	0.720
magic.gamma	0.732	0.648	0.498	0.662	0.874	0.812	0.827	0.831	0.874	0.746	0.828	0.881	0.803	**0.889**	0.858	0.860
mammography	0.861	0.909	0.520	0.755	0.934	0.932	0.927	0.925	0.945	0.856	0.934	**0.948**	0.864	0.946	0.923	0.934
mnist	0.803	0.846	0.523	0.705	0.959	**0.984**	**0.984**	0.983	0.969	0.967	0.965	0.982	0.846	0.980	0.968	0.979
musk	0.999	0.952	0.550	0.891	0.996	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.752	**1.000**	**1.000**	**1.000**
optdigits	0.674	0.612	0.522	0.385	0.997	0.996	**1.000**	**1.000**	0.998	0.987	**1.000**	0.999	0.777	**1.000**	0.998	0.999
PageBlocks	0.894	0.913	0.522	0.775	**0.959**	0.929	0.885	0.897	0.935	0.901	0.948	0.949	0.905	0.933	0.925	0.933
pendigits	0.955	0.910	0.485	0.718	0.999	0.996	0.997	0.997	0.998	0.993	0.996	**1.000**	0.923	0.999	0.997	0.993
satellite	0.699	0.750	0.503	0.739	0.932	0.806	0.856	0.849	0.903	0.867	0.866	0.928	0.836	**0.937**	0.935	0.914
satimage-2	0.992	0.966	0.525	0.971	0.992	0.992	**0.993**	**0.993**	0.969	0.959	0.977	0.986	0.912	0.988	0.969	0.974
shuttle	**0.996**	0.995	0.508	0.757	0.992	0.987	0.979	0.979	0.983	0.987	0.987	0.984	0.982	0.980	0.984	0.978
skin	0.684	0.391	0.500	0.499	0.998	0.903	0.951	0.955	0.995	0.999	0.983	0.998	0.999	**1.000**	0.996	0.976
SpamBase	0.633	0.660	0.500	0.544	0.887	0.889	0.913	0.931	0.938	0.868	0.886	0.946	0.807	0.945	0.928	**0.953**
speech	0.498	0.510	0.549	0.482	0.559	0.611	0.713	**0.740**	0.624	0.643	0.598	0.694	0.584	0.685	0.678	0.734
thyroid	0.981	0.979	0.521	0.918	0.986	0.996	**0.997**	**0.997**	0.994	0.983	0.996	0.996	0.879	**0.997**	0.993	**0.997**
vowels	0.765	0.440	0.407	0.798	0.871	0.971	0.970	0.979	0.977	0.953	0.961	0.995	0.837	**0.996**	0.976	0.991
Waveform	0.693	0.723	0.495	0.546	0.811	0.898	**0.909**	0.903	0.880	0.899	0.765	0.885	0.700	0.839	0.885	0.885
Wilt	0.427	0.395	0.502	0.479	0.960	0.646	0.689	0.696	0.954	0.981	0.876	0.970	0.922	**0.991**	0.732	0.976
yeast	0.382	0.387	0.477	0.458	0.644	0.459	0.663	0.666	0.646	0.703	0.663	0.672	0.660	0.699	**0.705**	0.650

Table A9: AUC-PR results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 5\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.036	0.036	0.042	0.038	**0.058**	0.041	0.042	0.045	0.046	0.043	0.040	0.048	0.037	0.040	0.035	0.036
annthyroid	0.336	0.260	0.078	0.359	0.402	0.400	0.425	0.458	0.541	0.482	0.609	0.602	0.363	0.575	0.658	**0.799**
Cardiotocography	0.441	0.436	0.254	0.345	0.602	0.760	0.767	**0.776**	0.708	0.610	0.652	0.720	0.433	0.736	0.649	0.691
fault	0.418	0.317	0.355	0.477	0.531	0.543	0.557	**0.565**	0.514	0.501	0.527	0.525	0.425	0.521	0.495	0.538
http	0.808	0.158	0.022	0.648	0.659	0.833	0.869	0.845	0.982	0.003	0.847	0.898	0.805	**1.000**	0.867	**1.000**
landsat	0.199	0.262	0.210	0.210	0.617	0.322	0.448	0.447	0.572	0.502	0.575	0.537	0.407	0.497	**0.682**	0.633
letter	0.098	0.076	0.165	0.141	0.176	0.148	0.158	0.159	0.201	0.134	0.109	0.137	**0.342**	0.169	0.118	0.126
magic.gamma	0.648	0.551	0.357	0.449	0.709	0.730	0.725	0.755	0.771	0.535	0.732	0.774	0.548	0.750	0.703	**0.792**
mammography	0.195	0.414	0.055	0.127	0.510	0.592	**0.610**	0.603	0.427	0.418	0.545	0.546	0.372	0.469	0.528	0.589
mnist	0.286	0.329	0.127	0.196	0.544	0.717	**0.784**	0.738	0.617	0.670	0.643	0.760	0.286	0.496	0.644	0.554
musk	0.970	0.343	0.105	0.464	0.597	0.653	**1.000**	**1.000**	**1.000**	0.961	0.895	**1.000**	0.246	**1.000**	0.974	0.951
optdigits	0.047	0.035	0.032	0.028	0.510	0.951	**0.991**	0.989	0.966	0.708	0.961	0.987	0.260	0.960	0.811	0.625
PageBlocks	0.469	0.517	0.146	0.346	0.714	0.678	0.654	0.679	0.642	0.539	**0.719**	0.680	0.456	0.696	0.602	0.638
pendigits	0.283	0.216	0.032	0.185	0.716	0.955	0.928	0.922	0.952	0.760	0.954	**0.980**	0.367	0.970	0.890	0.908
satellite	0.662	0.662	0.323	0.666	0.782	0.795	0.828	0.825	0.768	0.663	0.763	0.861	0.628	**0.869**	0.805	0.838
satimage-2	0.913	0.629	0.034	0.523	0.671	0.899	0.908	0.912	0.918	0.560	0.912	**0.930**	0.161	0.880	0.885	0.790
shuttle	**0.975**	0.957	0.081	0.464	0.954	0.974	0.968	0.967	0.962	0.956	0.967	0.966	0.959	0.966	0.953	0.960
skin	0.257	0.156	0.204	0.223	0.950	0.555	0.660	0.673	0.866	**0.994**	0.832	0.905	0.984	0.950	0.963	0.793
SpamBase	0.496	0.528	0.400	0.409	0.584	0.774	0.846	0.863	0.836	0.657	0.754	0.806	0.550	0.748	0.813	**0.889**
speech	0.021	0.019	0.049	0.017	0.024	0.045	0.052	0.057	0.042	0.044	0.027	**0.060**	0.040	0.046	0.037	0.054
thyroid	0.574	0.526	0.068	0.466	0.464	0.753	0.880	**0.888**	0.814	0.487	0.581	0.852	0.108	0.777	0.737	0.814
vowels	0.193	0.039	0.105	0.244	0.150	0.345	0.384	0.431	0.438	0.289	0.283	**0.602**	0.336	0.354	0.336	0.386
Waveform	0.063	0.064	0.032	0.043	0.180	0.121	0.135	0.177	0.182	**0.216**	0.156	0.181	0.141	0.154	0.182	0.190
Wilt	0.043	0.042	0.054	0.045	0.188	0.065	0.087	0.089	0.381	0.545	0.390	0.494	0.380	0.582	0.082	**0.633**
yeast	0.293	0.305	0.333	0.314	0.348	0.338	0.437	0.439	0.436	0.445	0.429	0.438	0.391	**0.467**	0.409	0.447

Table A10: AUC-PR results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 10\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.036	0.036	0.042	0.039	**0.068**	0.052	0.047	0.052	0.049	0.044	0.041	0.046	0.036	0.037	0.037	0.042
annthyroid	0.336	0.260	0.078	0.327	0.506	0.437	0.459	0.476	0.660	0.615	0.574	0.729	0.591	0.674	**0.835**	0.833
Cardiotocography	0.441	0.436	0.254	0.347	0.708	0.799	0.805	**0.817**	0.784	0.676	0.710	0.779	0.501	0.799	0.695	0.755
fault	0.418	0.317	0.355	0.480	0.560	0.577	0.586	**0.596**	0.542	0.521	0.543	0.537	0.490	0.575	0.538	0.567
http	0.808	0.158	0.022	0.451	0.829	0.928	0.891	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.801	**1.000**	**1.000**	**1.000**
landsat	0.199	0.262	0.210	0.215	0.704	0.357	0.454	0.486	0.632	0.568	0.516	0.597	0.474	0.614	**0.719**	0.708
letter	0.098	0.076	0.165	0.142	0.183	0.209	0.153	0.170	0.225	0.167	0.140	0.144	**0.305**	0.190	0.139	0.142
magic.gamma	0.648	0.551	0.357	0.473	0.754	0.748	0.718	0.748	0.817	0.594	0.758	0.822	0.622	0.819	0.735	**0.839**
mammography	0.195	0.414	0.055	0.127	0.562	0.606	**0.621**	0.606	0.515	0.370	0.520	0.566	0.475	0.573	0.565	0.544
mnist	0.286	0.329	0.127	0.198	0.678	**0.847**	0.846	0.820	0.752	0.737	0.737	0.844	0.443	0.781	0.724	0.799
musk	0.970	0.343	0.105	0.733	0.835	0.886	**1.000**	**1.000**	**1.000**	0.972	**1.000**	**1.000**	0.405	**1.000**	0.983	0.988
optdigits	0.047	0.035	0.032	0.027	0.811	0.973	0.994	0.991	0.979	0.713	0.961	0.992	0.323	**0.995**	0.885	0.746
PageBlocks	0.469	0.517	0.146	0.349	0.747	0.719	0.645	0.662	0.666	0.616	**0.774**	0.695	0.526	0.695	0.689	0.711
pendigits	0.283	0.216	0.032	0.186	0.893	0.963	0.914	0.914	0.955	0.881	0.946	**0.985**	0.427	0.980	0.946	0.865
satellite	0.662	0.662	0.323	0.653	0.829	0.789	0.830	0.828	0.846	0.749	0.772	**0.877**	0.636	0.872	0.829	0.851
satimage-2	0.913	0.629	0.034	0.527	0.868	0.910	0.897	0.886	0.911	0.851	0.907	**0.919**	0.525	0.916	0.873	0.877
shuttle	**0.975**	0.957	0.081	0.363	0.964	0.971	0.968	0.967	0.966	0.971	0.970	0.971	0.967	0.968	0.961	0.965
skin	0.257	0.156	0.204	0.212	0.971	0.532	0.658	0.675	0.937	**0.989**	0.945	0.959	0.982	0.944	0.970	0.904
SpamBase	0.496	0.528	0.400	0.411	0.701	0.802	0.857	0.879	0.884	0.745	0.834	0.886	0.635	0.868	0.862	**0.934**
speech	0.021	0.019	0.049	0.017	0.025	0.056	0.065	0.068	**0.085**	0.066	0.044	0.065	0.064	0.067	0.061	0.072
thyroid	0.574	0.526	0.068	0.475	0.606	0.870	**0.903**	0.883	0.870	0.750	0.810	0.882	0.368	0.895	0.840	0.893
vowels	0.193	0.039	0.105	0.247	0.192	0.477	0.646	0.705	0.676	0.506	0.501	**0.770**	0.410	0.676	0.418	0.711
Waveform	0.063	0.064	0.032	0.043	**0.263**	0.146	0.160	0.189	0.221	0.245	0.195	0.248	0.137	0.218	0.232	0.220
Wilt	0.043	0.042	0.054	0.046	0.387	0.077	0.087	0.089	0.436	0.663	0.462	0.611	0.492	**0.831**	0.090	0.752
yeast	0.293	0.305	0.333	0.313	0.392	0.349	0.434	0.440	0.443	**0.489**	0.462	0.439	0.429	0.486	0.472	0.443

Table A11: AUC-PR results of model comparison on 25 real-world datasets w.r.t. $\gamma_l = 20\%$.

Dataset	Typical						MLP			AutoEncoder			ResNet		Transformer	
	Iforest	ECOD	Deep SVDD	GAN omaly	Deep SAD	REPEN	DevNet	PReNet	MLP-Overlap	FEAWAD (Sup)	FEAWAD (Weak)	AE-Overlap	ResNet	ResNet-Overlap	FTT	FTT-Overlap
ALOI	0.036	0.036	0.042	0.038	**0.069**	0.046	0.047	0.045	0.045	0.051	0.047	0.044	0.040	0.052	0.041	0.039
annthyroid	0.336	0.260	0.078	0.372	0.624	0.441	0.457	0.478	0.677	0.703	0.622	0.739	0.704	0.748	**0.858**	0.818
Cardiotocography	0.441	0.436	0.254	0.354	0.793	0.830	**0.838**	0.837	0.826	0.743	0.768	0.835	0.600	**0.838**	0.794	0.828
fault	0.418	0.317	0.355	0.485	0.582	**0.610**	0.581	0.604	0.510	0.517	0.570	0.520	0.536	0.592	0.548	0.565
http	0.808	0.158	0.022	0.451	0.891	0.928	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
landsat	0.199	0.262	0.210	0.228	**0.763**	0.389	0.528	0.487	0.688	0.695	0.563	0.523	0.587	0.680	0.753	0.745
letter	0.098	0.076	0.165	0.143	0.197	0.276	0.216	0.260	0.239	0.206	0.195	0.175	**0.344**	0.268	0.197	0.211
magic.gamma	0.648	0.551	0.357	0.518	0.799	0.746	0.720	0.743	0.817	0.674	0.726	0.826	0.684	**0.850**	0.767	0.819
mammography	0.195	0.414	0.055	0.119	0.600	0.617	0.614	0.613	0.513	0.477	0.618	0.600	0.510	0.602	0.601	**0.629**
mnist	0.286	0.329	0.127	0.195	0.810	**0.908**	0.871	0.888	0.850	0.808	0.786	0.883	0.598	0.902	0.798	0.872
musk	0.970	0.343	0.105	0.752	0.981	0.994	**1.000**	**1.000**	**1.000**	0.999	**1.000**	**1.000**	0.592	**1.000**	**1.000**	0.999
optdigits	0.047	0.035	0.032	0.028	0.961	0.989	0.996	0.994	0.987	0.887	0.993	0.992	0.517	**0.999**	0.967	0.982
PageBlocks	0.469	0.517	0.146	0.395	**0.787**	0.724	0.672	0.692	0.696	0.725	0.785	0.752	0.720	0.710	0.708	0.738
pendigits	0.283	0.216	0.032	0.188	0.984	0.973	0.933	0.940	0.982	0.919	0.965	**0.990**	0.733	0.986	0.969	0.958
satellite	0.662	0.662	0.323	0.646	0.874	0.790	0.832	0.829	0.854	0.794	0.801	0.888	0.741	**0.903**	0.870	0.882
satimage-2	0.913	0.629	0.034	0.535	0.906	0.923	**0.928**	0.919	0.926	0.877	0.926	0.920	0.752	0.927	0.907	0.911
shuttle	**0.975**	0.957	0.081	0.415	0.967	0.972	0.967	0.967	0.966	0.973	0.971	0.972	0.968	0.969	0.964	0.969
skin	0.257	0.156	0.204	0.202	0.981	0.513	0.658	0.679	0.954	0.988	0.871	0.988	0.990	**0.998**	0.974	0.857
SpamBase	0.496	0.528	0.400	0.416	0.811	0.838	0.858	0.885	0.905	0.814	0.854	0.923	0.759	0.928	0.879	**0.938**
speech	0.021	0.019	0.049	0.017	0.027	**0.160**	0.101	0.117	0.084	0.075	0.074	0.070	0.119	0.108	0.063	0.104
thyroid	0.574	0.526	0.068	0.475	0.785	0.866	0.911	0.898	0.859	0.751	0.871	0.897	0.574	0.891	**0.916**	0.914
vowels	0.193	0.039	0.105	0.253	0.341	0.833	0.807	0.838	0.825	0.654	0.803	0.921	0.621	0.925	0.757	**0.941**
Waveform	0.063	0.064	0.032	0.043	**0.335**	0.185	0.217	0.227	0.282	0.286	0.224	0.310	0.192	0.282	0.328	0.265
Wilt	0.043	0.042	0.054	0.053	0.573	0.087	0.086	0.088	0.476	0.799	0.537	0.579	0.590	**0.874**	0.110	0.822
yeast	0.293	0.305	0.333	0.318	0.445	0.349	0.462	0.467	0.438	**0.535**	0.477	0.476	0.493	0.509	0.514	0.453

.5 Additional Results of AD Loss Function Exploration

In addition to the main paper that demonstrates the embedding variations on the vowels dataset in Section 4.3.1, here we provide another example of the skin dataset, as shown in Figure A1. Compared to the other loss functions, our proposed Overlap loss better retains the ringlike shape in the embedding of input feature while achieving satisfactory detection performance.

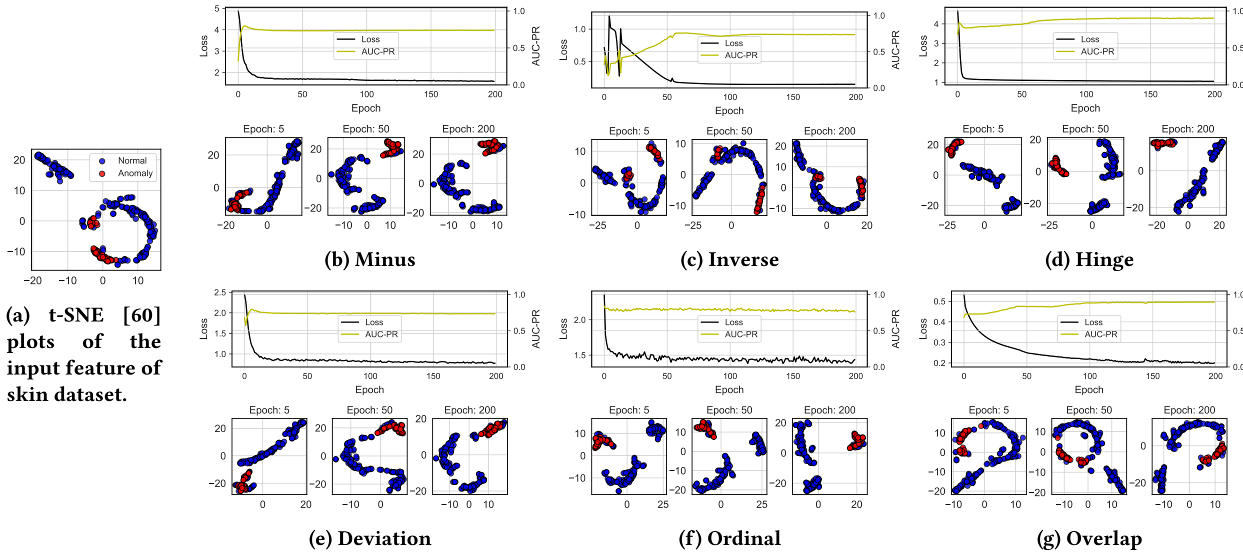
We follow [21, 58] to generate the following four types of synthetic anomalies, which are further used to evaluate different AD loss functions. The AUC-ROC results of loss function comparison on different types of anomalies also indicate that our proposed Overlap loss significantly outperforms other counterparts, as is shown in Table A12.

- **Local anomalies** refer to the anomalies deviant from their local neighborhoods [9]. GMM procedure [43, 58] is used to generate synthetic normal samples, and then scale the covariance matrix $\hat{\Sigma} = \alpha\hat{\Sigma}$ by a scaling parameter $\alpha = 5$ to generate local anomalies.
- **Global anomalies** are generated from a uniform distribution $\text{Unif}\left(\alpha \cdot \min\left(\boldsymbol{x}^{\boldsymbol{k}}\right), \alpha \cdot \max\left(\boldsymbol{x}^{\boldsymbol{k}}\right)\right)$, where the boundaries are defined as the *min* and *max* of an input feature, e.g., k-th feature $\boldsymbol{x}^{\boldsymbol{k}}$, and $\alpha = 1.1$ controls the outlyingness of anomalies.
- **Dependency anomalies** refer to the samples that do not follow the dependency structure that normal data follows [42], i.e., the input features of dependency anomalies are assumed to be independent of each other. Vine Copula [1] method is applied to model the dependency structure of original data, where the probability density function of generated anomalies is set to complete independence by removing the modeled dependency (see [42]). KDE method estimates the probability density function of features and generates normal samples.
- **Clustered anomalies**, also known as group anomalies [32], exhibit similar characteristics [15, 36]. We scale the mean feature vector of normal samples by $\alpha = 5$, i.e., $\hat{\boldsymbol{\mu}} = \alpha\hat{\boldsymbol{\mu}}$, where α controls the distance between anomaly clusters and the normals, and use the scaled GMM to generate anomalies.

Table A12: Loss function comparison on different types of anomalies generated based on the 25 real-world datasets.

Loss	Local	Global	Clustered	Dependency
Minus	0.629	0.936	0.996	0.738
Inverse	0.547	0.823	0.937	0.570
Hinge	0.607	0.938	0.997	0.761
Deviation	0.588	0.959	0.990	0.652
Ordinal	0.604	0.954	0.994	0.687
Overlap	**0.742**	**0.981**	**0.998**	**0.847**

We further investigate two case studies by generating visualized two-dimensional synthetic samples of the above local and clustered anomalies, as shown in Figure A2. The anomaly ratios of these two datasets are set to 5%. The results indicate that all the compared loss functions can correctly detect anomalies for the two-dimensional clustered anomalies (with 1.000 AUC-ROC and AUC-PR). This result can be expected since few labeled clustered anomalies can already represent similar behaviors of the entire clustered anomalies. For the local anomalies, however, we observe most of the compared loss functions perform poorly. In contrast, Overlap loss achieves better detection performance, and successfully learns a suitable decision boundary (see Figure A2l), where the learned decision boundary fits well with the local anomalies that are often overlapped or surrounded by the normal samples.



(a) t-SNE [60] plots of the input feature of skin dataset.

(b) Minus

(c) Inverse

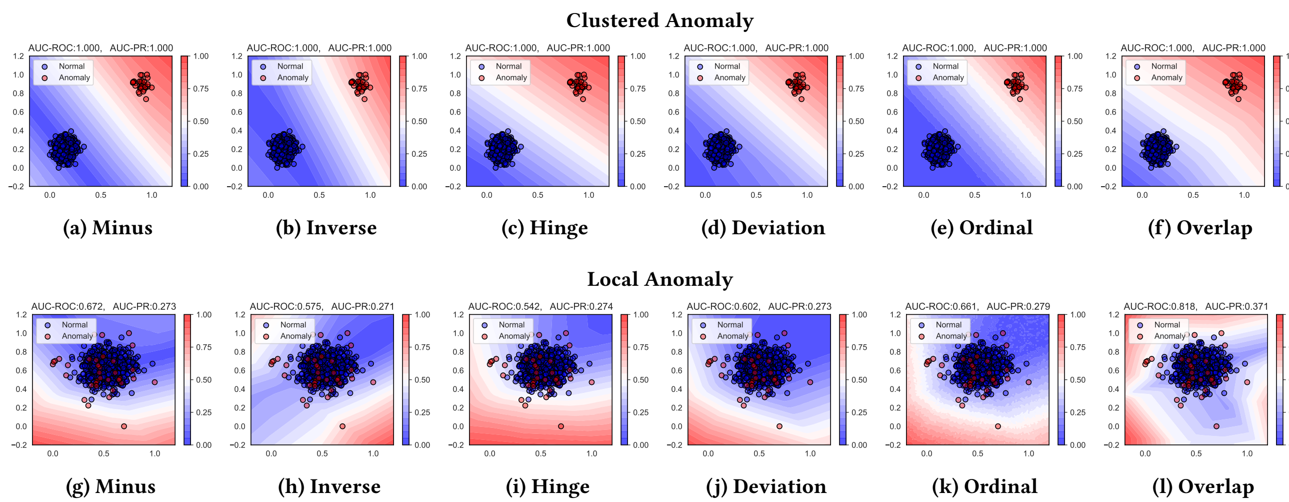
(d) Hinge

(e) Deviation

(f) Ordinal

(g) Overlap

Figure A1: Training loss along with the AUC-PR performance on testing set of different loss function based AD models, where the skin dataset is specified for comparison. The transformed embeddings of the input feature are demonstrated, which corresponds to 5, 50, and 200 training epochs, respectively.



(a) Minus (b) Inverse (c) Hinge (d) Deviation (e) Ordinal (f) Overlap

(g) Minus (h) Inverse (i) Hinge (j) Deviation (k) Ordinal (l) Overlap

Figure A2: Decision boundaries of different loss functions on the local anomalies. The output anomaly scores are normalized to $[0, 1]$ for comparison. Both AUC-ROC and AUC-PR performances are displayed in the title above each subfigure.