Multi-Sensor Object Anomaly Detection: Unifying Appearance, Geometry, and Internal Properties

Supplementary Material

A. Data Collection Details

As described in Sec 3.3 of the main text, the duration of thermal stimulation in Lock-in infrared thermography depends on the material properties and size of the objects. Objects with higher density and larger volume generally require a longer duration. This duration is controlled by two primary parameters: the lock-in period and the lock-in frequency. The lock-in period is responsible for the number of thermal stimulation, and the lock-in frequency determines the time intervals between each stimulation. Table A provides detailed information on the materials and dimensions of 15 objects, along with their respective lock-in periods and frequencies.

Table A. Properties and Thermal Stimulation Parameters of Objects

Category	Dimensions [mm]			Material	Lock-in Period	Lock-in Frequency [Hz]
	Length	*Width*	*Height*			
Capsule	20.0	8.0	8.0	Gelatin	30	1.0
Cotton	80.0	80.0	1.0	Fibre	30	1.0
Cube	100.0	100.0	10.0	Plastic	30	0.2
Piggy	45.0	30.0	35.0	Plastic	30	1.0
Screen	130.0	60.0	0.3	Glass	30	1.0
Flat pad	16.0	16.0	0.5	Metal	40	1.0
Screw	15.0	12.0	12.0	Metal	60	1.0
Nut	15.0	15.0	5.0	Metal	60	1.0
Spring pad	12.0	12.0	2.0	Metal	40	1.0
Button cell	12.0	12.0	5.0	Metal	50	1.0
Toothbrush	17.5	10.0	15.0	Plastic	30	2.0
Zipper	250.0	26.0	1.5	Fibre + Metal	30	1.0
Light	35.0	22.0	22.0	Plastic + Metal	90	0.5
Plastic cylinder	30.0	30.0	10.0	Nylon	60	1.0
Solar panel	40.0	40.0	3.0	Silicon	30	0.2

B. More Dataset Samples

Due to page limit, only a few dataset samples are shown in the main text. To provide a more intuitive view of our dataset, below are additional dataset samples(Fig. A–I). Note that the first row represents the RGB image, the second row represents the infrared image, and the third row represents the point cloud. Each column represents a normal or abnormal sample.

C. Implementation details

We employ two Transformer-based feature extractors to independently extract features from RGB/Infrared and point cloud data. For RGB/Infrared feature extraction, we use the ViT-B/8 model, which is pretrained on ImageNet with DINO. This model processes images resized to 224 × 224 pixels and outputs 784 patch features per image. For point cloud feature extraction, we use the PointMAE, pretrained on the ShapeNet dataset. Outputs from layers 3, 7, 11 are used to represent our 3D features. During training, we apply the AdamW optimizer with a learning rate set to 0.001, running the model for 200 epochs. All experiments are conducted on

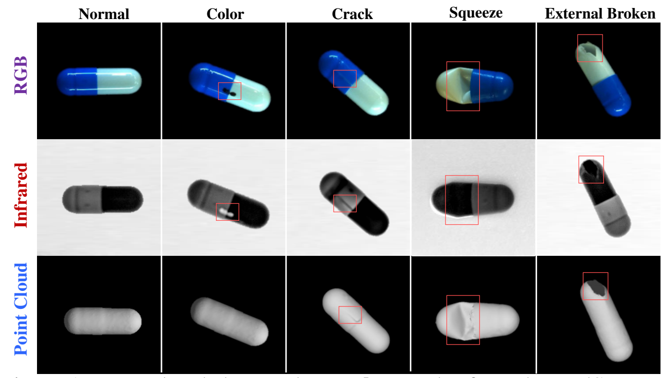

Figure A. Normal and abnormal **capsule** samples from the MulSen-AD Dataset.

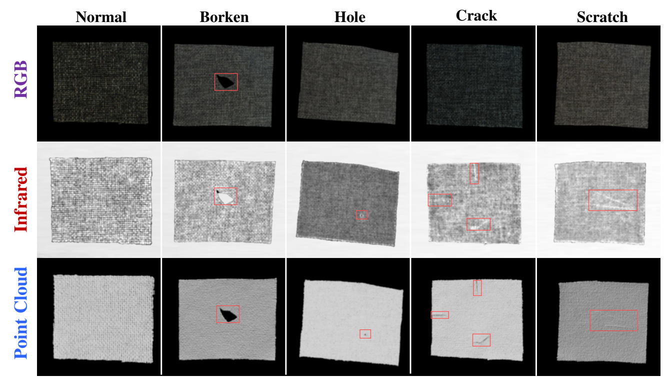

Figure B. Normal and abnormal **cotton** samples from the MulSen-AD Dataset.

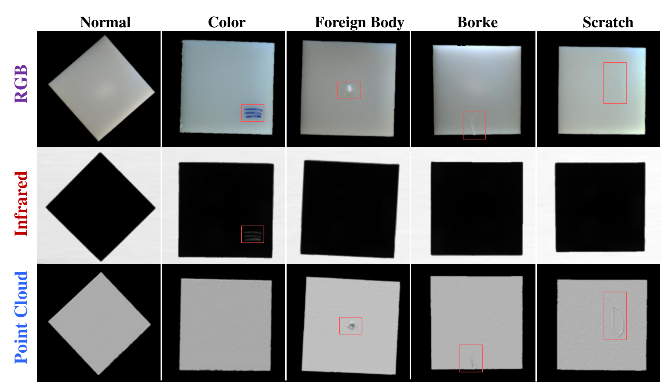

Figure C. Normal and abnormal **cube** samples from the MulSen-AD Dataset.

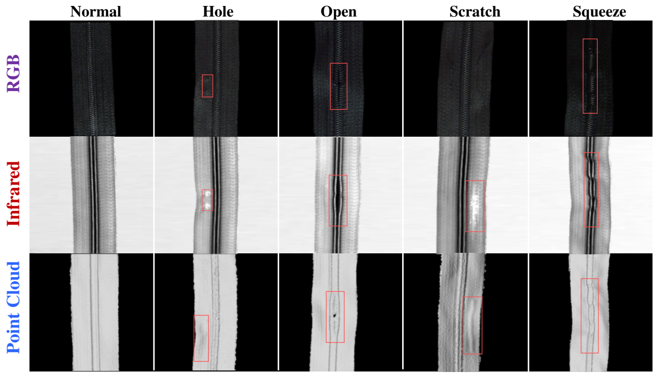

Figure D. Normal and abnormal **zipper** samples from the MulSen-AD Dataset.

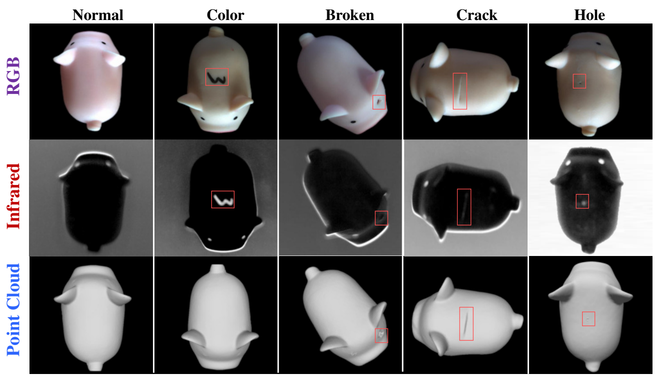

Figure E. Normal and abnormal **piggy** samples from the MulSen-AD Dataset.

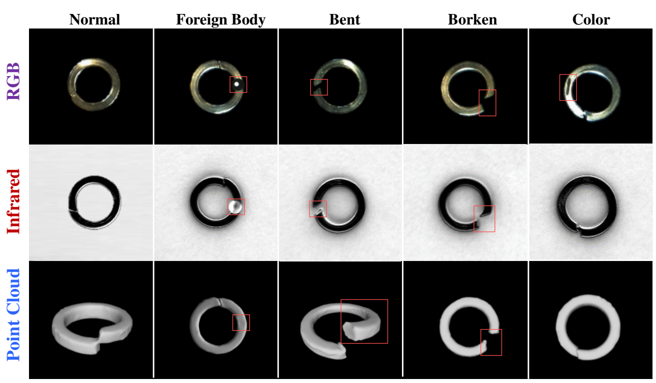

Figure F. Normal and abnormal **spring pad** samples from the MulSen-AD Dataset.

a Tesla V100 GPU.

D. Single 3D Benchmark

Due to the page limit, we only give the MulSen-AD Benchmark in the main text. Here we show the Single 3D Benchmark, including object-level Auroc in Table B, point-level Auroc in Table C. In the MulSen-AD setting, an object is labeled as abnormal if any one of the three modalities (RGB images, infrared images, or point clouds) is labeled as abnormal. **However, in the 3D-AD setting, an object is labeled as abnormal only if the point cloud specifically is labeled as abnormal.**

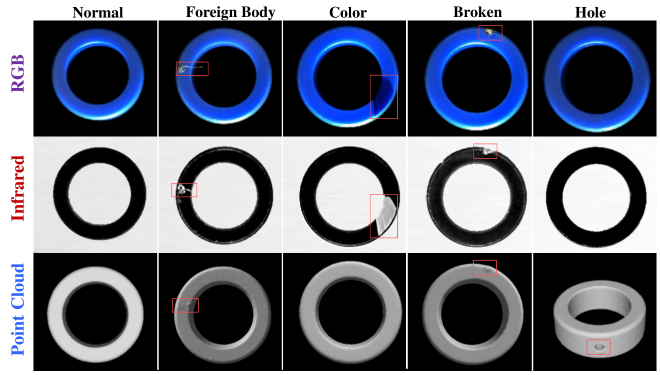

Figure G. Normal and abnormal **plastic cylinder** samples from the MulSen-AD Dataset.

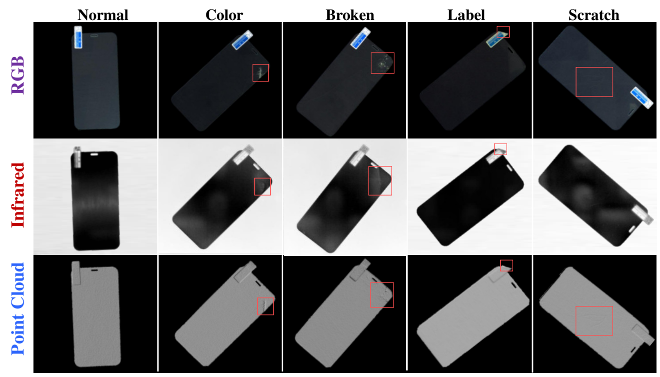

Figure H. Normal and abnormal **screen** samples from the MulSen-AD Dataset.

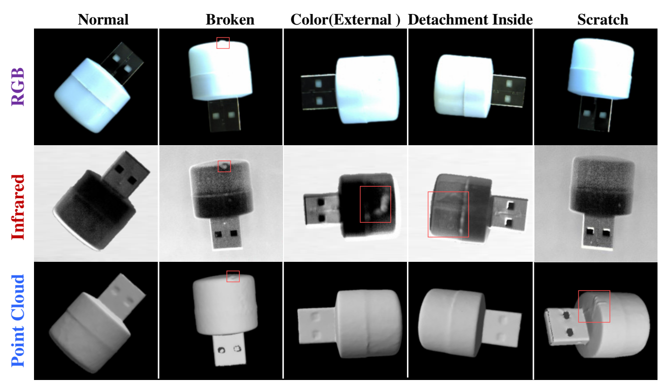

Figure I. Normal and abnormal **light** samples from the MulSen-AD Dataset.

E. Ablation Study on Memory Bank and Decision Unit Choices

This section presents the ablation study results for the memory bank and decision unit components. Specifically, we compare different choices for both components and their

Table B. SINGEBENCH-3D for MulSen-AD dataset. The score indicates object-level AUROC ↑. The best result of each category is highlighted in bold.

Category	BTF		M3DM		PatchCore			IMRNet	Reg3D-AD
	Raw	*FPFH*	*PointMAE*	*PointBERT*	*FPFH*	*FPFH+Raw*	*PointMAE*		
Capsule	0.641	0.874	0.731	0.671	0.898	0.905	0.903	0.601	**0.912**
Cotton	0.775	0.320	0.568	**0.805**	0.253	0.263	0.197	0.585	0.430
Cube	0.603	0.655	0.463	0.458	**0.723**	0.668	0.722	0.432	0.569
Spring pad	0.764	0.872	0.698	0.517	0.986	**1.000**	0.965	0.651	0.951
Screw	0.764	0.872	0.663	0.955	0.979	0.931	**0.997**	0.742	0.972
Screen	0.584	0.788	0.906	0.928	0.916	**0.950**	0.897	0.378	0.641
Piggy	0.818	0.831	0.164	0.447	**1.000**	0.997	0.982	0.729	0.866
Nut	0.789	0.883	0.783	0.751	0.971	**0.989**	**0.989**	0.812	0.797
Flat pad	0.698	0.918	0.885	0.772	**1.000**	0.893	0.944	0.714	0.908
Plastic cylinder	0.728	0.866	0.462	0.706	**0.941**	0.908	0.936	0.621	0.765
Zipper	0.505	0.662	0.701	0.698	0.797	**0.813**	0.739	0.630	0.470
Button cell	0.567	0.500	0.549	0.659	**0.915**	0.687	0.797	0.702	0.782
Toothbrush	0.882	0.562	0.803	0.901	**0.905**	0.888	0.891	0.615	0.812
Solar panel	0.474	0.531	0.385	0.395	0.624	0.605	0.612	0.344	**0.660**
Light	0.903	0.859	0.579	0.653	0.975	**1.000**	0.992	0.457	0.897
Mean	0.711	0.721	0.628	0.705	**0.860**	0.833	0.840	0.601	0.749

Table C. SINGEBENCH-3D for MulSen-AD dataset. The score indicates point-level AUROC ↑. The best result of each category is highlighted in bold.

Category	BTF		M3DM		PatchCore			IMRNet	Reg3D-AD
	Raw	*FPFH*	*PointMAE*	*PointBERT*	*FPFH*	*FPFH+Raw*	*PointMAE*		
Capsule	0.639	0.917	0.777	0.753	0.917	0.919	**0.921**	0.423	0.877
Cotton	0.412	0.581	0.663	**0.699**	0.554	0.546	0.528	0.507	0.521
Cube	0.441	**0.803**	0.613	0.710	0.575	0.437	0.417	0.566	0.626
Spring pad	0.659	0.780	0.568	0.652	0.629	0.601	0.621	0.401	**0.802**
Screw	0.577	0.582	0.453	0.443	0.578	**0.610**	0.597	0.456	0.540
Screen	0.469	**0.612**	0.529	0.567	0.609	0.587	0.532	0.352	0.466
Piggy	0.735	**0.871**	0.617	0.572	0.848	0.624	0.603	0.512	0.635
Nut	0.640	**0.924**	0.631	0.687	0.903	0.896	0.897	0.369	0.807
Flat pad	0.604	**0.715**	0.626	0.583	0.707	0.678	0.630	0.542	0.692
Plastic cylinder	0.662	**0.858**	0.510	0.652	0.830	0.766	0.769	0.412	0.670
Zipper	0.390	0.532	0.496	**0.563**	0.552	0.545	0.502	0.496	0.536
Button cell	0.671	0.694	0.797	**0.799**	0.382	0.512	0.478	0.485	0.706
Toothbrush	0.471	**0.634**	0.501	0.386	0.605	0.604	0.606	0.519	0.472
Solar panel	0.536	**0.727**	0.539	0.601	0.202	0.265	0.274	0.533	0.609
Light	0.665	**0.710**	0.480	0.495	0.707	0.706	0.696	0.415	0.651
Mean	0.571	**0.729**	0.587	0.611	0.640	0.620	0.605	0.467	0.641

Memory Bank	Decision Unit	AUROC(↑)
GMM	OC-SVM	0.795
PatchCore	KNN	0.951
PatchCore	LOF	0.939
PatchCore	IsolationForest	0.744
PatchCore	**OC-SVM (Ours)**	**0.961**

Table D. Ablation study results for memory bank and decision unit, measured by object-level AUROC.

impact on the performance. For the memory bank, we compare PatchCore's method with traditional Gaussian Mixture Model (GMM) modeling. In the decision unit section, we evaluate three different algorithms: K-Nearest Neighbors (KNN), Local Outlier Factor (LOF), and Isolation Forest (IF).The results, summarized in Table D, highlight the significance of selecting the appropriate methods for both components. Notably, **our approach, combining PatchCore with OC-SVM, delivers the best performance.**

Potential negative social impacts. Our dataset was collected with permission from the factory, so no negative social impact will exist.