Supplementary Materials of “Anomaly Heterogeneity Learning for Open-set Supervised Anomaly Detection”

Jiawen Zhu[1], Choubo Ding[2], Yu Tian[3], and Guansong Pang[1*]

[1]School of Computing and Information Systems, Singapore Management University
[2]Australian Institute for Machine Learning, University of Adelaide
[3]Harvard Ophthalmology AI Lab, Harvard University

A. Dataset Details

A.1. Key Data Statistics

We conduct extensive experiments on nine real-world Anomaly Detection (AD) datasets. Table 1 provides key data statistics for all the datasets used in this study. We follow exactly the same settings as prior Open-set Supervised Anomaly Detection (OSAD) studies [4, 7]. Particularly, we follow the original settings of MVTec AD and split the normal samples into training and test sets; for the other eight datasets, the normal samples are randomly split into training and test sets using a 3:1 proportion.

Dataset			Original Training	Original Test	
	$\vert C\vert$	Type	Normal	Normal	Anomaly
Carpet	5	Texture	280	28	89
Grid	5	Texture	264	21	57
Leather	5	Texture	245	32	92
Tile	5	Texture	230	33	83
Wood	5	Texture	247	19	60
Bottle	3	Object	209	20	63
Capsule	5	Object	219	23	109
Pill	7	Object	267	26	141
Transistor	4	Object	213	60	40
Zipper	7	Object	240	32	119
Cable	8	Object	224	58	92
Hazelnut	4	Object	391	40	70
Metal_nut	4	Object	220	22	93
Screw	5	Object	320	41	119
Toothbrush	1	Object	60	12	30
MVTec AD	73	-	3,629	467	1,258
AITEX	12	Texture	1,692	564	183
SDD	1	Texture	594	286	54
ELPV	2	Texture	1,131	377	715
Optical	1	Object	10,500	3,500	2,100
Mastcam	11	Object	9,302	426	451
BrainMRI	1	Medical	73	25	155
HeadCT	1	Medical	75	25	100
Hyper-Kvasir	4	Medical	2,021	674	757

Table 1. Statistical details for nine real-world AD datasets, with the first 15 rows displaying detailed information for subsets of the MVTec AD dataset.

MVTec AD [1] is a widely-used dataset that enables researchers to benchmark the performance of anomaly detection methods in the context of industrial inspection applications. The dataset includes over 5,000 images that are divided into 15 object and texture categories. Each category contains a training set of anomaly-free images, as well as a test set that includes images with both defects and defect-free images.

AITEX [9] is a textile fabric database that comprises 245 images of 7 different fabrics, including 140 defect-free images (20 for each type of fabric) and 105 images with various types of defects.

SDD [10] is a collection of images captured in a controlled industrial environment, using defective production items as the subject. The dataset includes 52 images with visible defects and 347 product images without any defects.

ELPV [3] is a collection of 2,624 high-resolution grayscale images of solar cells extracted from photovoltaic modules. These images were extracted from 44 different solar modules, and include both intrinsic and extrinsic defects known to reduce the power efficiency of solar modules.

Optical [11] is a synthetic dataset created to simulate real-world industrial inspection tasks for defect detection. The dataset comprises ten individual subsets, with the first six subsets (referred to as development datasets) intended for algorithm development purposes. The remaining four subsets (known as competition datasets) can be used to evaluate algorithm performance.

Mastcam [5] is a novelty detection dataset constructed from geological images captured by a multispectral imaging system installed on Mars exploration rovers. The dataset comprises typical images as well as images of 11 novel geologic classes. Each image includes a shorter wavelength (color) channel and a longer wavelengths (grayscale) channel.

BrainMRI [8] is a dataset for brain tumor detection obtained from magnetic resonance imaging (MRI) of the brain.

HeadCT [8] is a dataset consisting of 100 normal head CT

[*] Corresponding author: G. Pang (gspang@smu.edu.sg)

slices and 100 slices with brain hemorrhage, without distinction between the types of hemorrhage. Each slice is from a different person, providing a diverse set of images for researchers to develop and test algorithms for hemorrhage detection and classification in medical imaging applications.

Hyper-Kvasir [2] is a large-scale open gastrointestinal dataset which is collected during real gastro- and colonoscopy procedures. It is comprised of four distinct parts, including labeled image data, unlabeled image data, segmented image data, and annotated video data.

B. Implementation Details

B.1. Generating Heterogeneous Anomaly Distribution Datasets

Our proposed approach creates a diverse collection of data subsets by randomly selecting a subset of normal clusters and incorporating labeled anomaly examples to form the support and query sets for learning heterogeneous anomaly distributions. Specifically, we generate each data subset as follows.

Normal Samples in Each Data Subset. We employ the K-means algorithm to cluster normal samples into three groups. In each data subset, we randomly select two clusters to form the support and query sets. We follow this way to generate six such data subsets. Furthermore, we include an additional subset consisting of all normal samples to capture a holistic view of the distribution of normal instances, in which normal samples are randomly divided into two parts to form the support and query sets.

Abnormal Samples in Each Data Subset. To simulate open-set environments in heterogeneous anomaly distribution generation component, each anomaly distribution D_i is splited into two disjoint subsets, *i.e.*, $\mathcal{D}_i = \{\mathcal{D}_i^s, \mathcal{D}_i^q\}$, which correspond to support and query sets respectively, with the support set $\mathcal{D}_i^s = \mathcal{X}_{n,i}^s \cup \mathcal{X}_{a,i}^s$ used to train our base model ϕ_i and the query set $\mathcal{D}_i^q = \mathcal{X}_{n,i}^q \cup \mathcal{X}_{a,i}^q$ used to validate its open-set performance. For the setting with $M = 10$, we randomly select 50% samples from seen anomaly set and one normal cluster to form support set, while the remaining 50% seen anomalies are used in the query set. Under the protocol of having only one seen anomaly example, the example is included in both sets.

To enhance the variety of anomaly samples in our approach, we introduce three distinct anomaly augmentation techniques to generate pseudo anomaly samples: CutMix [12], CutPaste [6], DREAM Mask [13]. These techniques are randomly applied to each heterogeneous anomaly distribution subset to introduce diverse types of pseudo anomalies.

Algorithm 1 Anomaly Heterogeneity Learning (AHL)

```
Input: Input $\mathcal{D} = \{\mathbf{x}, y\}, \{\phi\}_1^T, \psi$
Output: Output $g$
/* Heterogeneous Anomaly Distribution Generation */
Construct $\mathcal{D}_i$ through grouping training set $\mathcal{D}$ into $T$ groups
/* Collaborative Differentiable Learning */
for $epoch = 1$ to $N$ do
    Update parameter of base model $\phi_i$ for $\mathcal{D}_i$ based on Eq.(1)
    /* Learning Importance Scores of Individual Anomaly Distributions */
    if $epoch >= 5$ then
        Compute the generalization error $r_i$ and importance score $w_i$ for $\phi_i$ with the help of sequential model $\psi$ via Eq.(6) and Eq.(7), respectively
    else
        Treat all $\phi$ equally, $w_i$ = $\frac{1}{T}$
    end if
    Update parameter of $g$ based on Eq.(4)
    Set the parameters of $\phi_i$ as the new weight parameters of $g$
end for
```

B.2. The Algorithm of AHL

The overall objective of our AHL framework is to achieve a unified and robust AD model via synthesizing anomaly heterogeneities learned from various heterogeneous anomaly distributions. We summarize the Anomaly Heterogeneity Learning (AHL) procedure in Algorithm 1. Specifically, our framework first generates T heterogeneous anomaly datasets, with each subset is sampled from the training set and contains a mixture of normal samples and (pseudo) anomaly samples, denoted as $\{\mathcal{D}_i\}_{i=1}^T$. In doing so, each subset is characterized by different set of normality/abnormality patterns,embodying heterogeneous anomaly distributions. We then employ a set of base models, denoted as $\{\phi_i\}_{i=1}^T$, to learn the underlying anomalous heterogeneity from these heterogeneous anomaly distributions. Moreover, a self-supervised sequential modeling approach is introduced to estimate the generalization errors r_i and importance scores w_i for each base model. Finally, we incorporate knowledge learned from heterogeneous anomaly distributions into a unified heterogeneous abnormality detection model g to capture richer anomaly heterogeneity.

AHL is a generic framework, in which off-the-shelf open-set anomaly detectors can be easily plugged and gain significantly improved generalization and accuracy in detecting both seen and unseen anomalies. Once we choose a base anomaly detector, the training strategy and objective

Dataset	One Anomaly Examples (Random)							Ten Anomaly Examples (Random)						
	SAOE	MLEP	FLOS	DevNet	DRA	AHL(DevNet)	AHL(DRA)	SAOE	MLEP	FLOS	DevNet	DRA	AHL(DevNet)	AHL(DRA)
Carpet	0.766±0.098	0.701±0.091	0.755±0.026	0.778±0.055	**0.873±0.035**	0.802±0.018	**0.877±0.004**	0.755±0.136	0.781±0.049	0.780±0.009	0.864±0.012	**0.945±0.014**	0.867±0.006	**0.953±0.001**
Grid	0.921±0.032	0.839±0.028	0.871±0.076	0.868±0.031	**0.972±0.016**	0.872±0.032	**0.975±0.005**	0.952±0.011	0.980±0.009	0.966±0.005	0.901±0.016	**0.990±0.008**	0.914±0.003	**0.992±0.002**
Leather	0.996±0.007	0.781±0.020	0.791±0.057	0.874±0.016	**0.988±0.003**	**0.880±0.005**	**0.988±0.001**	1.000±0.000	0.813±0.158	0.993±0.004	0.986±0.033	**1.000±0.000**	**0.996±0.008**	**1.000±0.000**
Tile	0.935±0.034	0.927±0.036	0.787±0.038	0.872±0.035	**0.966±0.014**	0.909±0.007	**0.968±0.001**	0.944±0.013	**0.988±0.009**	0.952±0.010	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
Wood	0.948±0.009	0.660±0.142	0.927±0.065	0.917±0.029	**0.987±0.012**	**0.947±0.020**	**0.987±0.003**	0.976±0.031	**0.999±0.002**	**1.000±0.000**	**0.999±0.000**	0.998±0.013	**1.000±0.000**	0.998±0.000
Bottle	0.989±0.019	0.927±0.090	0.975±0.023	0.986±0.012	**1.000±0.000**	**0.994±0.004**	**1.000±0.000**	0.998±0.003	0.981±0.004	0.995±0.002	0.996±0.005	**1.000±0.000**	**0.998±0.001**	**1.000±0.000**
Capsule	0.611±0.109	0.558±0.075	0.666±0.020	0.567±0.042	**0.646±0.029**	0.581±0.202	**0.665±0.030**	0.850±0.054	0.818±0.063	0.902±0.017	0.872±0.017	**0.928±0.011**	0.885±0.012	**0.930±0.001**
Pill	0.652±0.078	0.656±0.061	0.745±0.064	0.779±0.018	**0.831±0.026**	0.781±0.087	**0.840±0.003**	0.872±0.049	0.845±0.048	**0.929±0.012**	0.882±0.008	**0.918±0.009**	0.900±0.004	**0.918±0.001**
Transistor	0.680±0.182	0.695±0.124	0.709±0.041	0.732±0.075	0.727±0.105	**0.737±0.098**	**0.796±0.003**	0.860±0.053	0.927±0.043	0.862±0.037	0.907±0.004	**0.919±0.003**	0.912±0.002	**0.926±0.009**
Zipper	0.970±0.033	0.856±0.086	0.885±0.033	0.914±0.027	**0.983±0.008**	0.928±0.006	**0.986±0.000**	0.995±0.004	0.965±0.002	0.990±0.008	0.992±0.008	**1.000±0.000**	**0.995±0.002**	**1.000±0.000**
Cable	0.819±0.060	0.688±0.017	0.790±0.039	0.790±0.086	**0.855±0.007**	0.793±0.091	**0.858±0.011**	0.862±0.022	0.857±0.062	0.890±0.063	0.901±0.006	**0.914±0.006**	0.907±0.004	**0.921±0.001**
Hazelnut	0.961±0.042	0.704±0.090	0.976±0.021	0.970±0.005	**0.982±0.005**	0.978±0.003	**0.989±0.004**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
Metal_nut	0.922±0.033	0.878±0.038	0.930±0.022	0.874±0.014	**0.950±0.011**	0.880±0.002	**0.952±0.003**	0.976±0.013	0.974±0.009	0.984±0.004	0.992±0.003	**0.997±0.005**	**0.997±0.003**	**0.998±0.000**
Screw	0.653±0.074	0.675±0.294	0.337±0.091	0.766±0.045	**0.902±0.032**	0.769±0.031	**0.927±0.009**	0.975±0.023	0.899±0.039	0.940±0.017	0.981±0.013	0.978±0.012	**0.984±0.005**	**0.985±0.002**
Toothbrush	0.686±0.110	0.617±0.058	0.731±0.028	**0.790±0.029**	0.675±0.019	**0.794±0.016**	0.710±0.007	0.865±0.062	0.783±0.048	0.900±0.008	**0.950±0.025**	0.908±0.007	**0.959±0.002**	0.921±0.007
MVTec AD	0.834±0.007	0.744±0.019	0.792±0.014	0.832±0.016	**0.889±0.013**	0.843±0.021	**0.901±0.003**	0.926±0.010	0.907±0.005	0.939±0.007	0.948±0.005	**0.966±0.002**	0.954±0.003	**0.970±0.002**
AITEX	0.675±0.094	0.564±0.055	0.538±0.073	0.609±0.054	0.693±0.031	**0.704±0.004**	**0.734±0.008**	0.874±0.024	0.867±0.037	0.841±0.049	0.889±0.007	0.892±0.007	**0.903±0.011**	**0.925±0.013**
SDD	0.781±0.009	0.811±0.045	0.840±0.043	0.851±0.003	**0.907±0.002**	0.864±0.001	**0.909±0.001**	0.955±0.020	0.783±0.013	0.967±0.018	0.985±0.004	**0.990±0.000**	**0.991±0.001**	**0.991±0.000**
ELPV	0.635±0.092	0.578±0.062	0.457±0.056	**0.810±0.024**	0.676±0.003	**0.828±0.005**	0.723±0.008	0.793±0.047	0.794±0.047	0.818±0.032	0.843±0.001	0.843±0.002	**0.849±0.003**	**0.850±0.004**
Optical	0.815±0.014	0.516±0.009	0.518±0.003	0.513±0.001	**0.880±0.002**	0.547±0.009	**0.888±0.007**	0.941±0.013	0.740±0.039	0.720±0.055	0.785±0.012	**0.966±0.002**	0.841±0.010	**0.976±0.004**
Mastcam	0.662±0.018	0.625±0.045	0.542±0.017	0.627±0.049	**0.709±0.011**	0.644±0.013	**0.743±0.003**	0.810±0.029	0.798±0.026	0.703±0.029	0.797±0.021	**0.849±0.003**	0.825±0.020	**0.855±0.005**
BrainMRI	0.531±0.060	0.632±0.017	0.693±0.036	**0.853±0.045**	0.747±0.001	**0.866±0.004**	0.760±0.013	0.900±0.041	0.959±0.011	0.955±0.011	0.951±0.007	**0.971±0.001**	0.959±0.008	**0.977±0.001**
HeadCT	0.597±0.022	0.758±0.038	0.698±0.092	0.755±0.029	**0.804±0.010**	0.781±0.007	**0.825±0.014**	0.935±0.021	0.972±0.014	0.971±0.004	**0.997±0.002**	0.988±0.001	**0.999±0.003**	0.993±0.002
Hyper-Kvasir	0.498±0.100	0.445±0.040	0.668±0.004	0.734±0.020	0.712±0.010	**0.768±0.015**	**0.742±0.015**	0.666±0.050	0.600±0.069	0.773±0.029	0.822±0.031	0.844±0.001	**0.873±0.009**	**0.880±0.003**

Table 2. AUC results(mean±std) on nine real-world AD datasets under the general setting. Best results and the second-best results are respectively highlighted in **Red** and **Bold**.

function should be consistent with it. Following the proposed loss of base models (*i.e.*, DRA and DevNet), we adopt the deviation loss [7] to evaluate the loss between predicted anomaly scores and ground truths in the whole training phase:

$$\begin{aligned}\ell_{dev}(\mathbf{x}, y; h) =& \mathbb{I}(y=0)|(dev(\mathbf{x};h))| \\ &+ \mathbb{I}(y=1)\max(0, m - dev(\mathbf{x};h)),\end{aligned}$$

where $\mathbb{I}(.)$ is an indicator function that is equal to one when the condition is true, and zero otherwise; $h(\cdot)$ denotes the anomaly detection model. $dev(\mathbf{x}) = \frac{h(\mathbf{x})-\mu_r}{\sigma_r}$ with μ_r and σ_r representing the mean and standard deviation of a set of sampled anomaly scores from the Gaussian prior distribution $\mathcal{N}(0,1)$. m is a confidence margin which defines a radius around the deviation.

C. Detailed Empirical Results

C.1. Full Results under General Setting

Table 2 shows the detailed comparison results of AHL and SOTA competing methods under the general setting. It includes the performance metrics of each category of MVTec AD dataset. Overall, our proposed AHL model consistently outperforms the baseline methods in both ten-shot and one-shot settings across all three application scenarios. AHL (DRA) achieves the best performance in terms of AUC. On average, AHL improves the AUC of DRA and DevNet by up to 4% and 9%, respectively.

C.2. Full Results under Hard Setting

To investigate the detection performance of AHL framework on novel anomaly classes, we evaluate its performance under the hard setting, and present the detailed results for six multi-subset datasets, including each anomaly class-level performance, in Table 3. Overall, our models – AHL (DRA) and AHL (DevNet) – achieve the best AUC results in both $M = 1$ and $M = 10$ setting protocols. Specifically, AHL improves the performances of DRA and DevNet by up to 3.2% and 3% AUC, respectively. The results here are consistent with the superiority performance of AHL in the general setting.

C.3. More Ablation Study Results

Class-level Results under Hard Setting. To evaluate the effectiveness of each module in our AHL approach (**+ HADG + CDL$^+$**), we compare it with the base model (**DRA**), the base model using randomly sampled data distribution subsets and initial CDL component (**+ CDL**), and the base model with HADG and initial CDL component (**+ HADG + CDL**). Table 4 shows the detailed results of class-level anomaly detection in the hard setting. The results show that all the modules in the AHL framework contribute to improving the detection performance on unseen anomaly classes, demonstrating the importance of anomaly heterogeneity.

Importance of Pseudo Anomalies. Moreover, since DevNet [7] does not use pseudo anomalies, we also evaluate the impact of removing the pseudo anomalies on the AHL framework when using DevNet as the base model. As shown in Figure 1, AHL (DevNet) remains substantially better than the original DevNet model in such cases.

Effectiveness of Data Overlapping in HADG Module. Table 5 shows the results of AHL (DRA) using three data overlapping types: overlap occurring exclusively in (1) the query set ($\mathcal{D}^s$->$\mathcal{D}^q$), (2) the support set ($\mathcal{D}^q$->$\mathcal{D}^s$), and (3) in both sets ($\mathcal{D}^s <> \mathcal{D}^q$). It is clear that the overlapping between $\mathcal{D}^q$ and $\mathcal{D}^s$ typically does not lead to performance improvement.

Dataset		One Example from One Anomaly Class							Ten Example from One Anomaly Class						
		SAOE	MLEP	FLOS	DevNet	DRA	AHL(DevNet)	AHL(DRA)	SAOE	MLEP	FLOS	DevNet	DRA	AHL(DevNet)	AHL(DRA)
Carpet	Color	0.763±0.100	0.547±0.056	0.467±0.278	0.701±0.046	**0.890±0.011**	0.718±0.009	**0.894±0.004**	0.467±0.067	0.698±0.025	0.760±0.005	0.774±0.009	**0.899±0.019**	0.778±0.004	**0.929±0.007**
	Cut	0.664±0.165	0.658±0.056	0.685±0.007	0.679±0.018	**0.890±0.024**	0.684±0.014	**0.934±0.003**	0.793±0.175	0.653±0.120	0.688±0.059	0.817±0.021	**0.942±0.012**	0.825±0.006	**0.943±0.002**
	Hole	0.772±0.071	0.653±0.065	0.594±0.142	0.729±0.032	**0.915±0.045**	0.736±0.062	**0.935±0.014**	0.831±0.125	0.674±0.076	0.733±0.014	0.808±0.016	**0.958±0.031**	0.815±0.036	**0.960±0.003**
	Metal	0.780±0.172	0.706±0.047	0.701±0.028	0.822±0.016	**0.877±0.013**	0.846±0.012	**0.931±0.007**	0.883±0.043	0.764±0.061	0.678±0.083	0.885±0.012	**0.916±0.017**	0.899±0.026	**0.921±0.003**
	Thread	0.787±0.204	0.831±0.117	0.941±0.005	0.937±0.017	**0.954±0.010**	0.941±0.006	**0.966±0.005**	0.834±0.297	0.967±0.006	0.946±0.005	0.981±0.005	**0.985±0.005**	0.984±0.013	**0.991±0.001**
	Mean	0.753±0.055	0.679±0.029	0.678±0.040	0.774±0.007	**0.905±0.006**	0.785±0.015	**0.932±0.003**	0.762±0.073	0.751±0.023	0.761±0.012	0.853±0.005	**0.940±0.006**	0.860±0.013	**0.949±0.002**
Metal_nut	Bent	0.864±0.032	0.743±0.013	0.851±0.046	0.817±0.033	**0.952±0.015**	0.831±0.020	**0.954±0.003**	0.901±0.023	0.956±0.013	0.827±0.075	0.907±0.018	**0.987±0.003**	0.909±0.016	**0.989±0.000**
	Color	0.857±0.037	0.835±0.075	0.821±0.059	0.903±0.019	**0.930±0.021**	0.910±0.008	**0.933±0.008**	0.879±0.018	0.945±0.039	0.978±0.008	**0.992±0.015**	0.956±0.009	**0.995±0.002**	0.958±0.001
	Flip	0.751±0.090	0.813±0.031	0.799±0.058	0.751±0.039	**0.931±0.017**	**0.755±0.022**	**0.931±0.001**	0.795±0.062	0.805±0.057	0.942±0.009	**0.982±0.010**	0.931±0.010	**0.987±0.003**	0.937±0.003
	Scratch	0.792±0.075	0.907±0.085	0.947±0.027	**0.974±0.061**	0.929±0.009	**0.981±0.035**	0.934±0.005	0.845±0.041	0.805±0.153	0.943±0.002	**0.998±0.005**	**0.998±0.006**	**0.998±0.001**	**0.999±0.00**
	Mean	0.816±0.029	0.825±0.023	0.855±0.024	0.861±0.019	**0.936±0.011**	0.869±0.004	**0.939±0.004**	0.855±0.016	0.878±0.058	0.922±0.014	0.970±0.009	0.968±0.006	**0.972±0.002**	**0.971±0.001**
AITEX	Broken end	**0.778±0.068**	0.441±0.111	0.645±0.030	0.702±0.037	0.696±0.057	**0.716±0.014**	0.704±0.005	0.712±0.068	**0.732±0.065**	0.585±0.037	0.658±0.062	0.708±0.062	0.688±0.013	**0.735±0.010**
	Broken pick	0.644±0.039	0.476±0.070	0.598±0.023	0.567±0.016	**0.719±0.004**	0.575±0.005	**0.727±0.003**	0.629±0.012	0.555±0.027	0.548±0.054	0.595±0.017	**0.671±0.034**	0.612±0.005	**0.683±0.002**
	Cut selvage	0.681±0.077	0.434±0.149	0.694±0.036	0.674±0.021	**0.751±0.006**	0.680±0.017	**0.753±0.007**	0.770±0.014	0.682±0.025	0.745±0.035	0.703±0.062	**0.777±0.021**	0.737±0.012	**0.781±0.006**
	Fuzzyball	**0.650±0.064**	0.525±0.157	0.525±0.043	0.629±0.103	0.631±0.018	0.644±0.031	**0.647±0.007**	**0.842±0.026**	0.677±0.223	0.550±0.082	0.736±0.101	0.749±0.033	0.755±0.002	**0.775±0.024**
	Nep	0.710±0.044	0.517±0.059	0.734±0.038	**0.741±0.011**	0.685±0.010	**0.754±0.012**	0.703±0.005	0.771±0.032	0.740±0.052	0.746±0.060	**0.806±0.039**	0.784±0.025	**0.836±0.007**	0.792±0.007
	Weft crack	0.582±0.108	0.400±0.029	0.546±0.114	0.561±0.085	**0.693±0.002**	0.588±0.018	**0.706±0.009**	0.618±0.172	0.370±0.037	0.636±0.051	0.614±0.097	**0.710±0.016**	0.624±0.005	**0.713±0.003**
	Mean	0.674±0.034	0.466±0.030	0.624±0.024	0.646±0.014	**0.696±0.011**	0.660±0.007	**0.707±0.007**	0.724±0.032	0.626±0.041	0.635±0.043	0.685±0.016	**0.733±0.011**	0.709±0.006	**0.747±0.002**
ELPV	Mono	0.563±0.102	0.649±0.027	0.717±0.025	0.620±0.057	**0.762±0.017**	0.638±0.019	**0.774±0.013**	0.569±0.035	0.756±0.045	0.629±0.072	0.639±0.067	**0.735±0.008**	0.663±0.007	**0.745±0.004**
	Poly	0.665±0.173	0.483±0.247	0.665±0.021	**0.705±0.011**	0.681±0.026	**0.717±0.007**	**0.705±0.006**	0.796±0.084	0.734±0.078	0.662±0.042	0.806±0.027	0.806±0.004	**0.842±0.003**	**0.831±0.011**
	Mean	0.614±0.048	0.566±0.111	0.691±0.008	0.663±0.008	**0.722±0.006**	0.678±0.006	**0.740±0.003**	0.683±0.047	0.745±0.020	0.646±0.032	0.722±0.018	**0.771±0.005**	0.752±0.005	**0.788±0.003**
Mastcam	Bedrock	0.636±0.072	0.532±0.036	0.499±0.056	0.508±0.107	**0.653±0.019**	0.533±0.065	**0.679±0.012**	0.636±0.068	0.512±0.062	0.499±0.098	0.586±0.012	**0.654±0.013**	0.589±0.010	**0.673±0.006**
	Broken-rock	**0.699±0.058**	0.544±0.088	0.569±0.025	0.558±0.016	0.640±0.023	0.572±0.014	**0.661±0.009**	**0.712±0.052**	0.651±0.063	0.608±0.085	0.562±0.033	0.704±0.007	0.572±0.024	**0.722±0.004**
	Drill-hole	**0.697±0.074**	0.636±0.066	0.539±0.077	0.555±0.026	0.642±0.035	0.563±0.012	**0.654±0.004**	0.682±0.042	0.660±0.002	0.601±0.009	0.590±0.074	**0.757±0.008**	0.610±0.075	**0.760±0.003**
	Drt	**0.735±0.020**	0.624±0.042	0.591±0.042	0.570±0.048	**0.733±0.027**	0.581±0.023	0.724±0.006	**0.761±0.062**	0.616±0.048	0.652±0.024	0.620±0.031	0.757±0.006	0.629±0.016	**0.772±0.004**
	Dump-pile	0.682±0.022	0.545±0.127	0.508±0.021	0.510±0.008	**0.741±0.022**	0.519±0.004	**0.756±0.011**	0.750±0.037	0.696±0.047	0.700±0.070	0.689±0.070	**0.757±0.008**	0.695±0.021	**0.802±0.005**
	Float	**0.711±0.041**	0.530±0.075	0.551±0.030	0.507±0.039	0.688±0.031	0.524±0.017	**0.702±0.005**	0.718±0.064	0.671±0.032	0.736±0.041	0.640±0.012	**0.749±0.009**	0.647±0.008	**0.765±0.002**
	Meteorite	**0.669±0.037**	0.476±0.014	0.462±0.077	0.436±0.033	0.604±0.020	0.463±0.008	**0.616±0.013**	0.647±0.030	0.473±0.047	0.568±0.053	0.561±0.053	**0.689±0.010**	0.572±0.015	**0.691±0.001**
	Scuff	**0.679±0.048**	0.492±0.037	0.508±0.070	0.496±0.121	0.573±0.017	0.515±0.013	**0.581±0.020**	**0.676±0.019**	0.504±0.052	0.575±0.042	0.447±0.043	0.626±0.016	0.506±0.104	**0.656±0.009**
	Veins	**0.688±0.069**	0.489±0.028	0.493±0.052	0.530±0.006	0.650±0.012	0.548±0.010	**0.687±0.017**	**0.686±0.053**	0.510±0.090	0.608±0.044	0.577±0.029	0.644±0.007	0.598±0.037	**0.650±0.003**
	Mean	**0.689±0.037**	0.541±0.007	0.524±0.013	0.519±0.014	0.658±0.021	0.535±0.003	**0.673±0.010**	0.697±0.014	0.588±0.016	0.616±0.021	0.588±0.025	**0.704±0.007**	0.602±0.008	**0.721±0.003**
Hyper-Kvasir	Barretts	0.382±0.117	0.438±0.111	0.703±0.040	0.682±0.007	**0.788±0.008**	0.701±0.013	**0.792±0.007**	0.698±0.037	0.540±0.014	0.764±0.066	**0.837±0.014**	0.820±0.005	**0.850±0.002**	0.829±0.002
	Barretts-short-seg	0.367±0.050	0.532±0.075	0.538±0.033	0.608±0.077	**0.643±0.013**	0.629±0.027	**0.651±0.006**	0.661±0.034	0.480±0.107	0.810±0.034	0.790±0.017	**0.829±0.006**	0.812±0.005	**0.895±0.003**
	Esophagitis-a	0.518±0.063	0.491±0.084	0.536±0.040	0.567±0.034	**0.759±0.015**	0.583±0.015	**0.760±0.006**	0.820±0.034	0.646±0.036	0.815±0.022	0.867±0.004	0.854±0.006	**0.876±0.002**	**0.878±0.021**
	Esophagitis-b-d	0.358±0.039	0.457±0.086	0.505±0.039	0.535±0.025	**0.604±0.022**	0.562±0.010	**0.622±0.014**	0.611±0.017	0.621±0.042	0.754±0.073	0.812±0.025	0.785±0.006	**0.842±0.003**	**0.815±0.010**
	Mean	0.406±0.018	0.480±0.044	0.571±0.004	0.598±0.006	**0.699±0.009**	0.619±0.005	**0.706±0.007**	0.698±0.021	0.571±0.014	0.786±0.021	0.827±0.008	0.822±0.013	**0.845±0.003**	**0.854±0.004**

Table 3. AUC results(mean±std) on nine real-world AD datasets under the hard setting. Best results and the second-best results are respectively highlighted in **Red** and **Bold**. Carpet and Meta_nut are two subsets of MVTec AD. The same set of datasets is used as in [4].

Datsset	Subset	DRA	+ CDL	+ HADG + CDL	+ HADG+ CDL$^+$
Carpet	Color	0.899±0.019	0.917±0.004	**0.919±0.003**	**0.929±0.007**
	Cut	0.942±0.012	0.938±0.004	**0.943±0.001**	**0.943±0.002**
	Hole	**0.958±0.031**	0.954±0.010	0.952±0.002	**0.960±0.003**
	Metal	0.916±0.017	**0.919±0.008**	0.914±0.006	**0.921±0.003**
	Thread	0.985±0.005	0.985±0.003	**0.988±0.002**	**0.991±0.001**
	Mean	0.940±0.006	**0.943±0.002**	**0.943±0.003**	**0.949±0.002**
AITEX	Broken end	0.708±0.062	0.714±0.011	**0.719±0.005**	**0.735±0.010**
	Broken pick	0.671±0.034	0.670±0.003	**0.678±0.005**	**0.683±0.002**
	Cut selvage	0.777±0.021	0.777±0.009	**0.779±0.018**	**0.781±0.006**
	Fuzzyball	0.749±0.033	0.742±0.010	**0.756±0.021**	**0.775±0.024**
	Nep	0.784±0.025	0.786±0.007	**0.788±0.005**	**0.792±0.007**
	Weft crack	0.710±0.016	0.708±0.003	**0.711±0.003**	**0.713±0.003**
	Mean	0.733±0.011	0.733±0.005	**0.739±0.007**	**0.747±0.002**
elpv	Mono	0.735±0.008	0.738±0.008	**0.739±0.006**	**0.745±0.004**
	Poly	0.806±0.004	0.809±0.006	**0.817±0.012**	**0.831±0.011**
	Mean	0.771±0.005	0.774±0.004	**0.784±0.004**	**0.788±0.003**
Hyper-Kvasir	Barretts	0.820±0.005	0.819±0.006	**0.822±0.009**	**0.829±0.002**
	Barretts-short-seg	0.829±0.006	0.864±0.012	**0.887±0.016**	**0.895±0.003**
	Esophagitis-a	0.854±0.006	0.863±0.000	**0.871±0.005**	**0.878±0.021**
	Esophagitis-b-d	0.785±0.006	0.795±0.005	**0.806±0.004**	**0.815±0.010**
	Mean	0.822±0.013	0.835±0.004	**0.847±0.008**	**0.854±0.004**

Table 4. Ablation study class-level results of AHL and its three main variants under hard settings. Best results and the second-best results are respectively highlighted in **Red** and **Bold**.

Dataset	$\mathcal{D}^s$->$\mathcal{D}^q$	$\mathcal{D}^q$->$\mathcal{D}^s$	$\mathcal{D}^s$ <> $\mathcal{D}^q$	Ours
AITEX	0.918	0.917	0.917	**0.925**
SDD	0.988	0.990	**0.991**	**0.991**
ELPV	0.853	0.844	0.842	**0.850**
BrainMRI	0.970	0.969	0.970	**0.977**
HeadCT	0.993	0.991	0.993	**0.993**
Hyper-Kvasir	0.872	0.875	0.876	**0.880**

Table 5. Results with various data overlapping in the support and query sets in HADG.

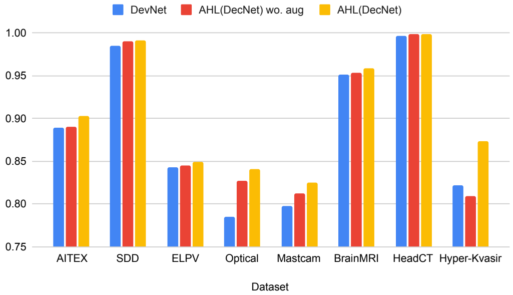

Figure 1. Comparison of the AUC performance between DevNet and AHL (DRA) without the pseudo anomaly augmentation generation module. Here *wo. aug* indicates the augmentation techniques are excluded during training.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1

[2] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Sigrun L Eskeland, Kristin Ranheim Randel, Konstantin Pogorelov, Mathias Lux, Duc Tien Dang Nguyen, et al. Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific data*, 7(1):283, 2020. 2

[3] Sergiu Deitsch, Vincent Christlein, Stephan Berger, Claudia Buerhop-Lutz, Andreas Maier, Florian Gallwitz, and Christian Riess. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*, 185:455–468, 2019. 1

[4] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7388–7398, 2022. 1, 4

[5] Hannah R Kerner, Kiri L Wagstaff, Brian D Bue, Danika F Wellington, Samantha Jacob, Paul Horton, James F Bell, Chiman Kwan, and Heni Ben Amor. Comparison of novelty detection methods for multispectral images in rover-based planetary exploration missions. *Data Mining and Knowledge Discovery*, 34:1642–1675, 2020. 1

[6] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021. 2

[7] Guansong Pang, Choubo Ding, Chunhua Shen, and Anton van den Hengel. Explainable deep few-shot anomaly detection with deviation networks. *arXiv preprint arXiv:2108.00462*, 2021. 1, 3

[8] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14902–14912, 2021. 1

[9] Javier Silvestre-Blanes, Teresa Albero-Albero, Ignacio Miralles, Rubén Pérez-Llorens, and Jorge Moreno. A public fabric database for defect detection methods and results. *Autex Research Journal*, 19(4):363–374, 2019. 1

[10] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3):759–776, 2020. 1

[11] Matthias Wieler and Tobias Hahn. Weakly supervised learning for industrial optical inspection. In *DAGM symposium in*, 2007. 1

[12] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 2

[13] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 2