Catching Both Gray and Black Swans: Open-set Supervised Anomaly Detection Supplementary Material*

Choubo Ding[1†], Guansong Pang[2†], Chunhua Shen[3]
[1]The University of Adelaide, [2]Singapore Management University, [3]Zhejiang University

A. Dataset Details

A.1. Key Statistics of Datasets

Tab. 1 summarizes the key statistics of these datasets. Below we introduce each dataset in detail. The normal samples in MVTec AD are split into training and test sets following the original settings. In other datasets, the normal samples are randomly split into training and test sets by a ratio of $3/1$.

MVTec AD [1] is a popular defect inspection benchmark that has 15 different classes, with each anomaly class containing one to several subclasses. In total the dataset contains 73 defective classes of fine-grained anomaly classes at the texture- or object-level.

AITEX [16] is a fabrics defect inspection dataset that has 12 defect classes, with pixel-level defect annotation. We crop the original 4096×256 image to several 256×256 patch image and relabel each patch by pixel-wise annotation.

SDD [17] is a defect inspection dataset images of defective production items with pixel-level defect annotation. We vertically and equally divide the original 500×1250 image into three segment images and relabel each image by pixel-wise annotation.

ELPV [3] is a solar cells defect inspection dataset in electroluminescence imagery. It contains two defect classes depending on solar cells: mono- and poly-crystalline.

Optical [20] is a synthetic dataset for defect detection on industrial optical inspection. The artificially generated data is similar to real-world tasks.

Mastcam [6] is a novelty detection dataset constructed from geological image taken by a multispectral imaging system installed in Mars exploration rovers. It contains typical images and images of 11 novel geologic classes. Images including shorter wavelength (color) channel and longer wavelengths (grayscale) channel and we focus on shorter wavelength channel in this work.

Table 1. Key Statistics of Image Datasets. The first 15 datasets compose the MVTec AD dataset.

Dataset	Original Training	Original Test		Anomaly Data	
	Normal	Normal	Anomaly	# Classes	Type
Carpet	280	28	89	5	Texture
Grid	264	21	57	5	Texture
Leather	245	32	92	5	Texture
Tile	230	33	84	5	Texture
Wood	247	19	60	5	Texture
Bottle	209	20	63	3	Object
Capsule	219	23	109	5	Object
Pill	267	26	141	7	Object
Transistor	213	60	40	4	Object
Zipper	240	32	119	7	Object
Cable	224	58	92	8	Object
Hazelnut	391	40	70	4	Object
Metal_nut	220	22	93	4	Object
Screw	320	41	119	5	Object
Toothbrush	60	12	30	1	Object
MVTec AD	3,629	467	1,258	73	-
AITEX	1,692	564	183	12	Texture
SDD	594	286	54	1	Texture
ELPV	1,131	377	715	2	Texture
Optical	10,500	3,500	2,100	1	Object
Mastcam	9,302	426	451	11	Object
BrainMRI	73	25	155	1	Medical
HeadCT	75	25	100	1	Medical
Hyper-Kvasir	2,021	674	757	4	Medical

BrainMRI [15] is a brain tumor detection dataset obtained by magnetic resonance imaging (MRI) of the brain.

HeadCT [15] is a brain hemorrhage detection dataset obtained by CT scan of head.

Hyper-Kvasir [2] is a large-scale open gastrointestinal dataset collected during real gastro- and colonoscopy procedures. It contains four main categories and 23 subcategories of gastro- and colonoscopy images. This work focuses on gastroscopy images with the anatomical landmark category as the normal samples and the pathological category as the anomalies.

To provide some intuitive understanding of what the anomalies and normal samples look like, we present some examples of normal and anomalous images for each dataset in Fig. 1.

*Corresponding author: CS (e-mail: chunhua@me.com). This work was in part done when GP and CS were with The University of Adelaide.
†First two authors contributed equally.

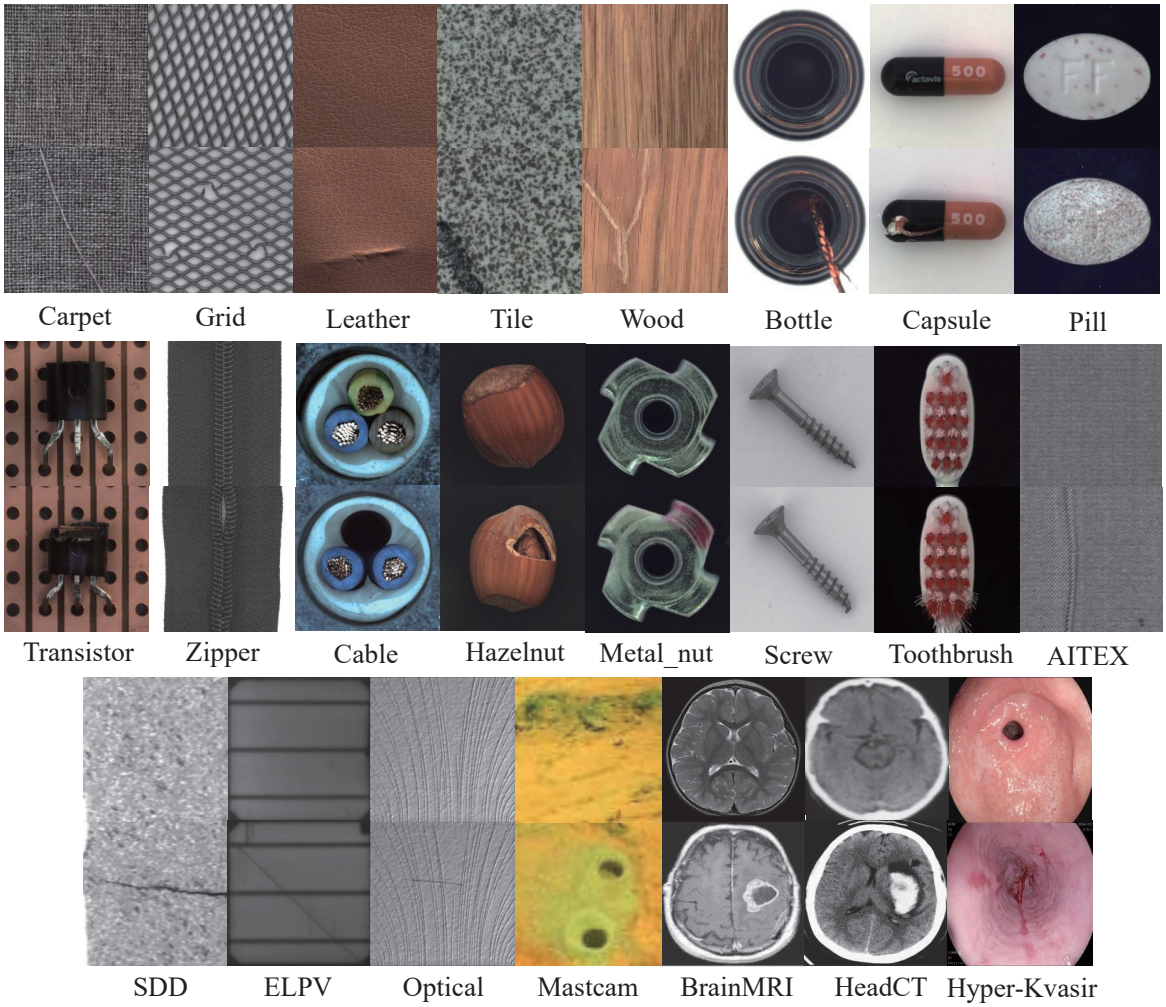

Figure 1. Examples of normal and anomalous images for each dataset. For each group of examples, the images on the top are normal, while the bottom ones are anomalous.

Table 2. Download Link of Image Datasets.

Dataset	Link
MVTec AD	https://tinyurl.com/mvtecad
AITEX	https://tinyurl.com/aitex-defect
SDD	https://tinyurl.com/KolektorSDD
ELPV	https://tinyurl.com/elpv-crack
Optical	https://tinyurl.com/optical-defect
Mastcam	https://tinyurl.com/mastcam
BrainMRI	https://tinyurl.com/brainMRI-tumor
HeadCT	https://tinyurl.com/headCT-tumor
Hyper-Kvasir	https://tinyurl.com/hyper-kvasir

A.2. Dataset Split

We have two experiment protocols, including general and hard settings. For the general setting, the few labeled anomaly samples are randomly drawn from all possible anomaly classes in the test set per dataset. These sampled anomalies are then removed from the test data. For the hard setting, the anomaly example sampling is limited to be drawn from one single anomaly class only, and all anomaly samples in this anomaly class are removed from the test set to ensure that the test set contains only unseen anomaly classes. As labeled anomalies are difficult to obtain due to their rareness and unknowingness, in both settings we use only very limited labeled anomalies, i.e., with the number of the given anomaly examples respectively fixed to one and ten.

Additionally, to have a cross analysis of the results in the one-shot and ten-shot scenarios, the one anomaly example in the one-shot scenarios is randomly sampled from the ten sampled anomaly examples in the ten-shot scenarios, and they are all evaluated on exactly the same test data – the test data used in ten-shot scenarios. That is, the only differ-

ence between the one-shot and ten-shot scenarios is on the training anomaly examples.

B. Implementation Details

In this section, we describe the implementation details of DRA and its competing methods.

B.1. Implementation of DRA

All input images are first resized to 448x448 or 224x224 according to the original resolution. We then use the ImageNet pre-trained ResNet-18 for the feature extraction network, which extracts a 512-dimensional feature map for an input image. This feature map is then fed to the subsequent abnormality/normality learning heads to learn disentangled abnormalities. The Patch-wise classifier in Plain Feature Learning adopted in the seen and pseudo abnormality learning heads is implemented by a 1x1 convolutional layer that yields the anomaly score of each vector in the feature map. The normality learning head utilizes a two-layer fully connected layer as the classifier, which first reduces a 512-dimensional feature vector to 256-dimensions and then yields anomaly scores. In the training phase, each input image is routed to different heads based on the labels, and each head computes the loss independently. All its heads are jointly trained using 30 epochs, with 20 iterations per epoch and a batch size of 48. Adam is used for the parameter optimization using an initial learning rate 10^{-3} with a weight decay of 10^{-2}.

In the inference phase, each head yields anomaly scores for the target image simultaneously. Since all heads are optimized by the same loss function, the anomaly scores generated by each head have the same semantic, and so we calculate the sum of all the anomaly scores (and a negated normal score) as the final anomaly score. In addition, to solve the multi-scale problem, we use an image pyramid module with two-layer image pyramid, which obtains anomaly scores at different scales by inputting original images of various sizes, and calculates the mean value as the final anomaly score.

For the reference sets in Latent Residual Abnormality Learning, we found that mixing some generated pseudo-anomaly samples into normal samples can further improve performance. We speculate that adding pseudo-anomaly samples can get more challenging residual samples to help the network adapt to extreme cases. Therefore, we use the dataset mixed with normal and pseudo-abnormal samples as the reference set in the final implementation.

B.1.1 Loss Function

DRA can be optimized using different anomaly score loss functions. We use deviation loss [11] in our final implementation to optimize DRA, because it is generally more effective and stable than other popular loss functions, as shown in the experimental results in Section C.2. Particularly, a deviation loss optimizes the anomaly scoring network by a Gaussian prior score, with the deviation specified as a Z-score:

$$dev(\mathbf{x}_i;\Theta) = \frac{g(f(\mathbf{x};\Theta_f);\Theta_g) - \mu_{\mathcal{R}}}{\sigma_{\mathcal{R}}}, \tag{1}$$

where $\mu_{\mathcal{R}}$ and $\sigma_{\mathcal{R}}$ is the mean and standard deviation of the prior-based anomaly score set drawn from $\mathcal{N}(\mu, \sigma^2)$. The deviation loss is specified using the contrastive loss [4] with the deviation plugged into:

$$\begin{aligned}\ell\big(\mathbf{x}_i, \mu_{\mathcal{R}}, \sigma_{\mathcal{R}}; \Theta\big) = (1-y_i)|dev(\mathbf{x}_i;\Theta)| \\ + y_i \max\big(0, a - dev(\mathbf{x}_i;\Theta)\big),\end{aligned} \tag{2}$$

where $y = 1$ indicate an anomaly and $y = 0$ indicate a normal sample, and a is equivalent to a Z-Score confidence interval parameter.

B.2. Implementation of Competing Methods

In the main text, we present five recent and closely related state-of-the-art (SOTA) competing methods. Here we introduce two additional competing methods. Following is the detailed description and implementation details of these seven methods:

KDAD [15] is an unsupervised deep anomaly detector based on multi-resolution knowledge distillation. We experiment with the code provided by its authors[1] and report the results. Since KDAD is unsupervised, it is trained with normal data only, but it is evaluated on exactly the same test data as DRA.

DevNet [11, 12] is a supervised deep anomaly detector based on a prior-based deviation. The results we report are based on the implementation provided by its authors[2].

FLOS [9] is a deep imbalanced classifier that learns a binary classification model using the class-imbalance-sensitive loss – focal loss. The implementation of FLOS is also taken from [11], which replaces the loss function of DevNet with the focal loss.

SAOE is a deep out-of-distribution detector that utilizes pseudo anomalies from both data augmentation-based and outlier exposure-based methods. Motivated by the success of using pseudo anomalies to improve anomaly detection in recent studies [7, 18], SAOE is implemented by learning both seen and pseudo abnormalities through a multi-class (*i.e.*, normal class, seen anomaly class, and pseudo anomaly class) classification head using the plain feature learning method as in DRA. In addition to this multi-class

[1] https://github.com/rohban-lab/Knowledge_Distillation_AD

[2] https://github.com/choubo/deviation-network-image

Table 3. AUC results (mean±std) of DRA and two additional competing methods under the general setting. All methods are trained using ten random anomaly examples, with the best results are **highlighted**.

Dataset	$\|\mathcal{C}\|$	DeepSAD	MINNS	DRA (Ours)
Carpet	5	0.791±0.011	0.876±0.015	**0.940**±0.027
Grid	5	0.854±0.028	0.983±0.016	**0.987**±0.009
Leather	5	0.833±0.014	0.993±0.007	**1.000**±0.000
Tile	5	0.888±0.010	0.980±0.003	**0.994**±0.006
Wood	5	0.781±0.001	**0.998**±0.004	**0.998**±0.001
Bottle	3	0.913±0.002	0.995±0.007	**1.000**±0.000
Capsule	5	0.476±0.022	0.905±0.013	**0.935**±0.022
Pill	7	0.875±0.063	**0.913**±0.021	0.904±0.024
Transistor	4	0.868±0.006	0.889±0.032	**0.915**±0.025
Zipper	7	0.974±0.005	0.981±0.011	**1.000**±0.000
Cable	8	0.696±0.016	0.842±0.012	**0.909**±0.011
Hazelnut	4	**1.000**±0.000	**1.000**±0.000	**1.000**±0.000
Metal_nut	4	0.860±0.053	0.984±0.002	**0.997**±0.002
Screw	5	0.774±0.081	0.932±0.035	**0.977**±0.009
Toothbrush	1	**0.885**±0.063	0.810±0.086	0.826±0.021
MVTec AD	-	0.830±0.009	0.939±0.011	**0.959**±0.003
AITEX	12	0.686±0.028	0.813±0.030	**0.893**±0.017
SDD	1	0.963±0.005	0.961±0.016	**0.991**±0.005
ELPV	2	0.722±0.053	0.788±0.028	**0.845**±0.013
Optical	1	0.558±0.012	0.774±0.047	**0.965**±0.006
Mastcam	11	0.707±0.011	0.803±0.031	**0.848**±0.008
BrainMRI	1	0.850±0.016	0.943±0.031	**0.970**±0.003
HeadCT	1	0.928±0.005	**0.984**±0.010	0.972±0.002
Hyper-Kvasir	4	0.719±0.032	0.647±0.051	**0.834**±0.004

classification, the outlier exposure module [5] in SAOE is implemented according to its authors[3], in which the MVTec AD [1] or LAG [8] dataset is used as external data. In all our experiments we removed the related data from the outlier data that has any overlapping with the target data to avoid data leakage.

MLEP [10] is a deep open set anomaly detector based on margin learning embedded prediction. The original MLEP[4] is designed for open set video anomaly detection, and we adapt it to image tasks by modifying the backbone network and training settings to be consistent with DRA.

Deep SAD [14] is a supervised deep anomaly detector that extends Deep SVDD [13] by using a few labeled anomalies and normal samples to learn more compact one-class descriptors. Particularly, it adds a new marginal constraint to the original Deep SVDD that enforces a large margin between labeled anomalies and the one-class center in latent space. The implementation of DeepSAD is taken from the original authors[5].

MINNS [19] is a deep multiple instance classification model, which is implemented based on [11].

[3] https://github.com/hendrycks/outlier-exposure
[4] https://github.com/svip-lab/MLEP
[5] https://github.com/lukasruff/Deep-SAD-PyTorch

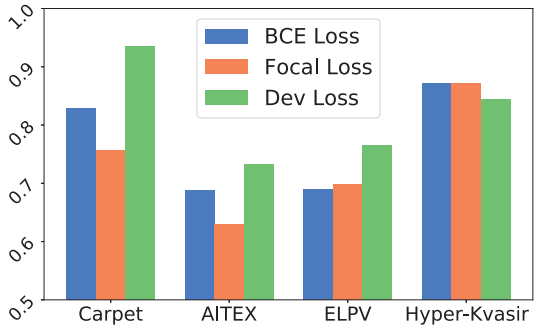

Figure 2. The AUC performance of our proposed method using different loss functions under the hard setting. We report the averaged results over all data subsets per dataset.

C. Additional Empirical Results

C.1. Additional Comparison Results

General Setting. We report the results of DRA and two additional competing methods under general setting in Tab 3. Our method achieves the best AUC performance in eight of the nine datasets and the close-to-best AUC performance in the another dataset. In the eight best-performing datasets, our method improves AUC by 2% to 19.1% over the best competing method.

Hard Setting. Tab. 4 shows the results of DRA and two additional competing methods under the hard setting. Our method performs best on most of the data subsets and achieves the best AUC performance on five of the six datasets at the dataset level. Our method improves from 9.2% to 24.4% over the suboptimal method in the other five datasets.

The experimental results in both settings show the superiority of our method compared to Deep SAD and MINNS.

C.2. Sensitivity w.r.t. Loss Function

In our paper, we use the deviation loss [11] in all our four heads by default. Here we vary the use of the loss function and analyze the impact of the loss function on the performance of anomaly detection. Any related binary classification loss functions may be used for training all the four heads of DRA. We evaluate the applicability of two additional popular loss functions, including binary cross-entropy loss and focal loss, in addition to deviation loss. The results are reported in Fig. 2, where all results are the averaged AUC of three independent runs of the experiments. In general, the deviation loss function, which is specifically designed for anomaly detection, has clear superiority on most cases. The two classification losses perform better on the medical dataset Hyper-Kvasir. Based on such empirical findings, the deviation loss function is generally recommended in DRA.

Table 4. AUC results of DRA and two additional competing methods under the hard setting, where models are trained with one known anomaly class and tested to detect the rest of all other anomaly classes. Each data subset is named by the known anomaly class.

Module		DeepSAD	MINNS	DRA (Ours)
Carpet	Color	0.736±0.007	0.767±0.011	**0.886**±0.042
	Cut	0.612±0.034	0.694±0.068	**0.922**±0.038
	Hole	0.576±0.036	0.766±0.007	**0.947**±0.016
	Metal	0.732±0.042	0.789±0.097	**0.933**±0.022
	Thread	0.979±0.000	0.982±0.008	**0.989**±0.004
	Mean	0.727±0.011	0.800±0.022	**0.935**±0.013
Metal_nut	Bent	0.821±0.023	0.868±0.033	**0.990**±0.003
	Color	0.707±0.028	**0.985**±0.018	0.967±0.011
	Flip	0.602±0.020	**1.000**±0.000	0.913±0.021
	Scratch	0.654±0.004	**0.978**±0.000	0.911±0.034
	Mean	0.696±0.012	**0.958**±0.008	0.945±0.017
AITEX	Broken_end	0.442±0.029	**0.708**±0.103	0.693±0.099
	Broken_pick	0.614±0.039	0.565±0.018	**0.760**±0.037
	Cut_selvage	0.523±0.032	0.734±0.012	**0.777**±0.036
	Fuzzyball	0.518±0.023	0.534±0.058	**0.701**±0.093
	Nep	0.733±0.017	0.707±0.059	**0.750**±0.038
	Weft_crack	0.510±0.058	0.544±0.183	**0.717**±0.072
	Mean	0.557±0.014	0.632±0.023	**0.733**±0.009
ELPV	Mono	0.554±0.063	0.557±0.010	**0.731**±0.021
	Poly	0.621±0.006	0.770±0.032	**0.800**±0.064
	Mean	0.588±0.021	0.663±0.015	**0.766**±0.029
Mastcam	Bedrock	0.474±0.038	0.419±0.025	**0.658**±0.021
	Broken-rock	0.497±0.054	**0.687**±0.015	0.649±0.047
	Drill-hole	0.494±0.013	0.651±0.035	**0.725**±0.005
	Drt	0.586±0.012	0.705±0.043	**0.760**±0.033
	Dump-pile	0.565±0.046	0.697±0.022	**0.748**±0.066
	Float	0.408±0.022	0.635±0.073	**0.744**±0.073
	Meteorite	0.489±0.010	0.551±0.018	**0.716**±0.004
	Scuff	0.502±0.010	0.502±0.040	**0.636**±0.086
	Veins	0.542±0.017	0.577±0.013	**0.620**±0.036
	Mean	0.506±0.009	0.603±0.016	**0.695**±0.004
Hyper-Kvasir	Barretts	0.672±0.017	0.679±0.009	**0.824**±0.006
	B.-short-seg	0.666±0.012	0.608±0.064	**0.835**±0.021
	Esophagitis-a	0.619±0.027	0.665±0.045	**0.881**±0.035
	E.-b-d	0.564±0.006	0.480±0.043	**0.837**±0.009
	Mean	0.630±0.009	0.608±0.014	**0.844**±0.009

C.3. Cross-domain Anomaly Detection

An interesting extension area of open-set anomaly detection is cross-domain anomaly detection, aiming at training detection models on a source domain to detect anomalies on datasets from a target domain different from the source domain. To demonstrate potential of our method in such setting, we report cross-domain AD results of our model DRA on all five texture anomaly datasets in MVTec AD in Tab. 5. DRA is trained on one of five datasets (source domain) and in fine-tuned with 10 epochs on the other four datasets (target domains) using normal samples only. The results show that the domain-adapted DRA significantly outperforms the SOTA unsupervised method KDAD that is directly trained on the target domain. This demonstrates some promising open-domain performance of DRA.

Table 5. AUC results of domain-adapted DRA and unsupervised method KDAD in texture datasets. The top row is the source domain and the left column is the target domain.

	carpet	grid	leather	tile	wood	KDAD
carpet	-	0.833	0.921	0.930	0.917	0.774
grid	0.983	-	0.924	0.940	0.916	0.749
leather	0.988	0.998	-	0.994	1.000	0.948
tile	0.917	0.971	0.958	-	0.955	0.911
wood	0.993	0.985	0.972	0.948	-	0.940

D. Failure Cases

Although DRA shows competitive results on most datasets, it still fails on individual datasets; the most notable is the toothbrush dataset. After in-depth research and analysis of the results, we believe the failure of the toothbrush dataset is mainly due to its small size of normal samples (60 normal samples, see Tab. 1). Due to the more complex architecture, DRA often requires a relatively larger set of normal training samples to learn the disentangled abnormalities, while simpler methods like FLOS and SAOE that perform mainly binary classification do not have this requirement and work better on this dataset. In practice, we need to pay attention to the available data size of the target task, and apply a lightweight network in DRA instead when facing small-scale tasks.

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad – a comprehensive real-world dataset for unsupervised anomaly detection. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, June 2019. 1, 4

[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):1–14, 2020. 1

[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] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, volume 2, pages 1735–1742, 2006. 3

[5] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *Proc. Int. Conf. Learn. Representations*, 2019. 4

[6] 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(6):1642–1675, 2020. 1

[7] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, pages 9664–9674, 2021. 3

[8] Liu Li, Mai Xu, Xiaofei Wang, Lai Jiang, and Hanruo Liu. Attention based glaucoma detection: A large-scale database and cnn model. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, June 2019. 4

[9] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 2980–2988, 2017. 3

[10] Wen Liu, Weixin Luo, Zhengxin Li, Peilin Zhao, Shenghua Gao, et al. Margin learning embedded prediction for video anomaly detection with a few anomalies. In *Proc. Int. Joint Conf. Artificial Intell.*, pages 3023–3030, 2019. 4

[11] 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. 3, 4

[12] Guansong Pang, Chunhua Shen, and Anton van den Hengel. Deep anomaly detection with deviation networks. In *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, pages 353–362, 2019. 3

[13] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proc. Int. Conf. Mach. Learn.*, pages 4393–4402, 2018. 4

[14] Lukas Ruff, Robert A Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. In *Proc. Int. Conf. Learn. Representations*, 2020. 4

[15] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad Hossein Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proc. IEEE Conf. Comp. Vis. Patt. Recogn.*, 2021. 1, 3

[16] 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

[17] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *Journal of Intelligent Manufacturing*, May 2019. 1

[18] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Proc. Advances in Neural Inf. Process. Syst.*, 33:11839–11852, 2020. 3

[19] Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu. Revisiting multiple instance neural networks. *Pattern Recognition*, 74:15–24, 2018. 4

[20] M Wieler and T Hahn. Weakly supervised learning for industrial optical inspection. In *DAGM Symposium*, 2007. 1