






GeneralAD: Anomaly Detection Across Domains by Attending to Distorted Features

Supplementary Material

Luc P.J. Sträter^{*} , Mohammadreza Salehi^{*} ,
Efstratios Gavves , Cees G. M. Snoek , and Yuki M. Asano 

University of Amsterdam, The Netherlands
lucstrater@gmail.com, s.salehidehnavi@uva.nl

A Datasets

Here we show a set of normal and anomaly samples that can be used in our experiments. As Figure 1 shows, we have conducted diverse and comprehensive experiments across different datasets to support the generality of our approach. Further details on the datasets are given below.

CIFAR [7]: CIFAR-10 and CIFAR-100 consist of 60,000 natural color images each, with a resolution of 32×32 . These images are divided into a training set of 50,000 and a testing set of 10,000. CIFAR-10 is structured into 10 equally sized classes, while CIFAR-100 is split into either 100 fine-grained or 20 coarse-grained classes, with our experiments following the coarse-grained classification.

Fashion MNIST [10]: FMNIST comprises 60,000 training samples and 10,000 test samples, each being a 28×28 grayscale image distributed among 10 unique classes.

View [4]: The View dataset is a collection of natural scene images divided into six classes, such as mountains and buildings. This dataset provides approximately $\sim 2,300$ images per class for training and ~ 500 images per class for testing.

Dogs vs. Cats [3]: This simple visualization dataset contains 25,000 images of dogs and cats.

Aircraft-FGVC [8]: This dataset comprises 10,200 images, each representing one of 102 different aircraft model variants, predominantly airplanes, with each variant having 100 images. An 80-20 train-test split is used. We evaluated the following ten classes: [91,96,59,19,37,45,90,68,74,89], aligning our approach with [9].

^{*} These authors contributed equally to this work.

Stanford-Cars [6] This dataset includes 16,185 images spread across 196 car classes. The dataset is divided into 8,144 training images and 8,041 testing images, with an approximately equal split for each class. We ran our experiments on the first 20 classes and tested on the entire test set, aligning our approach with [9].

MVTec-AD [2]: The MVTec Anomaly Detection dataset is recognized for its collection of 5354 high-resolution images from 15 different categories of industrial objects and textures, serving as the main testbed for anomaly detection algorithms in manufacturing. The defects present in these anomalies are varied, encompassing more than 70 different kinds, including scratches, dents, contaminations, and structural changes.

MVTec-LOCO [1]: This dataset targets anomaly localization with images of industrial products that contain logical and structural defects. It contains 3644 images from five different classes inspired by industrial inspection scenarios.

VisA [11]: The Visual Anomaly dataset is the largest industrial anomaly benchmark with images of various manufacturing anomalies. It contains 12 classes in 3 domains across 10,821 high-resolution images.

MPDD [5]: This smaller dataset is specifically designed to address the challenges of detecting defects in the fabrication of painted metal parts. It offers a realistic testing environment with variable spatial orientations, multiple objects, and non-homogeneous backgrounds, diverging from traditional lab-based AD datasets. It contains 1346 images across 6 categories.

B Extended Qualitative Results

We show more qualitative results of our method. Figure 2 shows the localization results of our method on two classes of MVTec AD [2]. Our method can accurately detect subtle defects in the dataset, such as in screw images. Figure 3 shows the localization results of our method on the Dogs vs. Cats dataset [3]. As can be seen, our method consistently focuses on discriminative features like the snout, whiskers, and ears.

C Per Class Results

Here we show the per-class results of our main results table. Table 1 shows the results for the semantic anomaly detection datasets, Table 2 for near anomaly detection datasets, and Table 3 for industrial anomaly detection datasets.

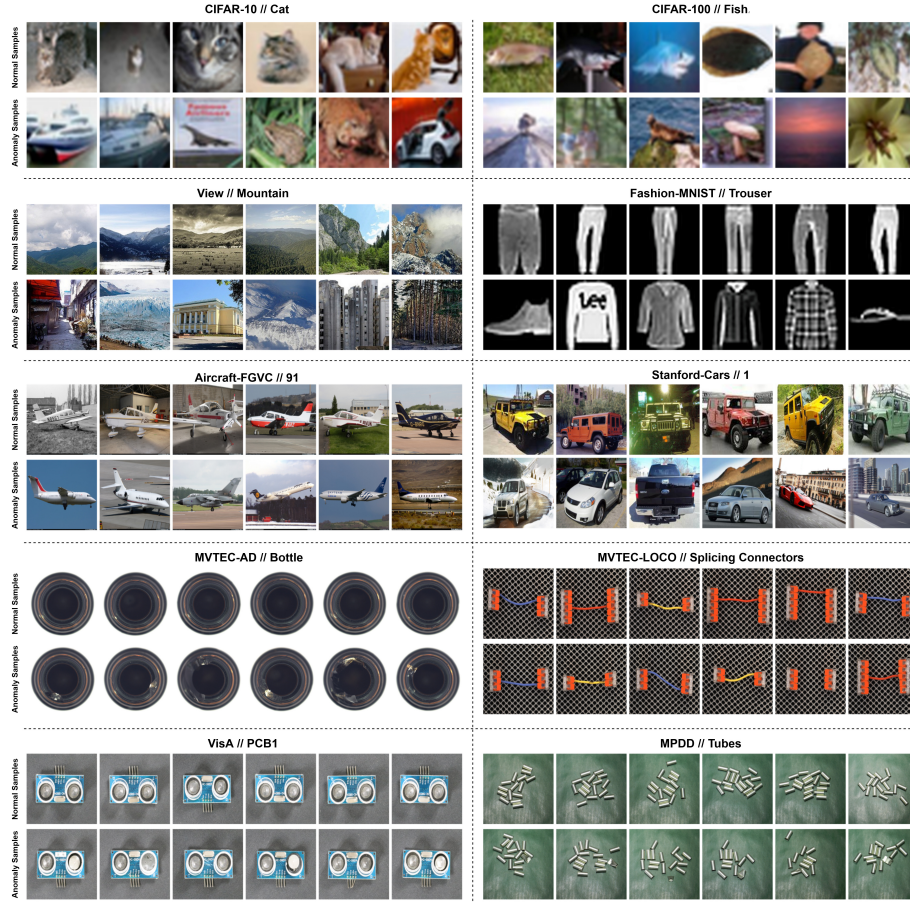


Fig. 1: Normal and Anomaly samples from all datasets.

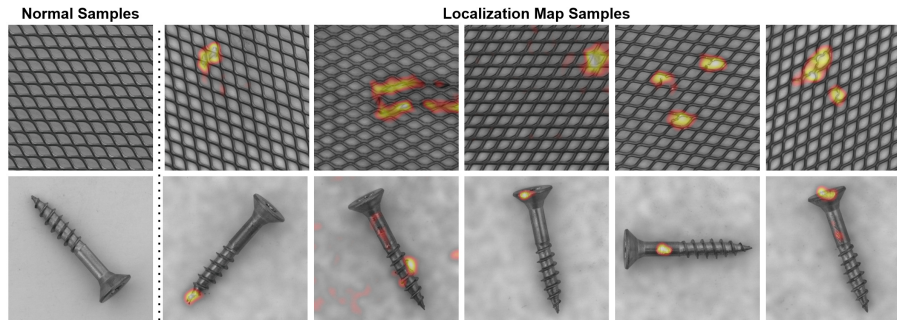


Fig. 2: Extended qualitative localization results GeneralAD on MVTEc-AD. We train the method on the normal samples specified in the first column, then specify the abnormality regions for each input pass at the test time.

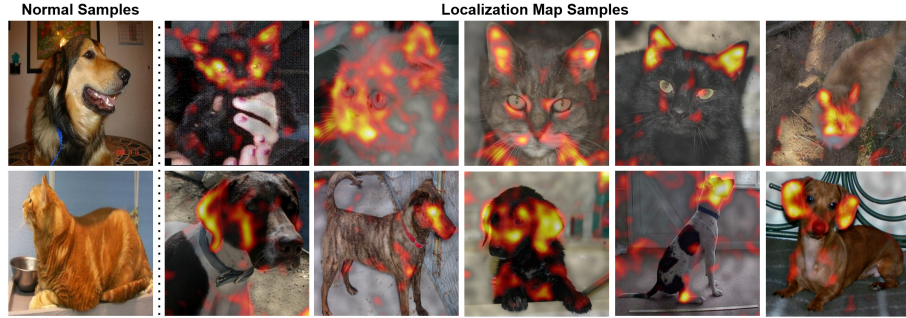


Fig.3: Extended qualitative localization results GeneralAD on Dogs vs. Cats. We train the method on the normal samples specified in the first column, then specify the abnormality regions for each input pass at the test time. As is shown, the method focuses mainly on the snout, whiskers, and ears to discriminate between dogs and cats.

Table 1: Image-level AUROC scores for GeneralAD on the semantic anomaly detection datasets. The per-class results of GeneralAD on the datasets CIFAR-10 and CIFAR-100 [7], FMNIST [10], and View [4].

CIFAR-10		CIFAR-100		FMNIST		View	
Class	I-AUROC (%)	Class	I-AUROC (%)	Class	I-AUROC (%)	Class	I-AUROC (%)
airplane	99.9	aquatic mammals	97.4	t-shirt/top	92.2	buildings	93.2
automobile	99.5	fish	98.1	trouser	99.0	forest	99.6
bird	99.5	flowers	99.7	pullover	93.1	glacier	94.5
cat	97.7	food containers	99.3	dress	95.2	mountain	94.5
deer	99.1	fruit and vegetables	99.5	coat	93.3	sea	95.7
dog	98.5	household electrical devices	98.2	sandal	97.9	street	98.1
frog	99.8	household furniture	99.5	shirt	85.4		
horse	99.5	insects	99.1	sneaker	98.3		
ship	99.8	large carnivores	97.9	bag	99.1		
truck	99.8	large man-made outdoor things	96.7	ankle boot	98.4		
		large natural outdoor scenes	98.0				
		large omnivores and herbivores	97.5				
		medium-sized mammals	96.2				
		non-insect invertebrates	98.6				
		people	99.4				
		reptiles	98.6				
		small mammals	98.1				
		trees	98.4				
		vehicles 1	98.7				
		vehicles 2	98.5				
All avg.	99.3		98.4		95.2		95.9

Table 2: Image-level AUROC scores for GeneralAD on the near anomaly detection datasets. The per-class results of GeneralAD on the datasets Aircraft-FGVC [8] and Stanford-Cars [6].

Aircraft-FGVC		Stanford-Cars	
Class	I-AUROC (%)	Class	I-AUROC (%)
91	98.7	1	98.5
96	99.5	2	76.6
59	90.5	3	90.4
19	91.8	4	86.9
37	99.4	5	87.9
45	91.8	6	83.9
90	96.4	7	76.2
68	90.7	8	82.7
74	92.8	9	78.6
89	94.5	10	91.5
		11	91.1
		12	86.1
		13	91.4
		14	88.5
		15	87.5
		16	97.5
		17	90.8
		18	89.9
		19	84.8
		20	86.0
All avg.	94.6		87.3

Table 3: Image-level AUROC scores for GeneralAD on the industrial anomaly detection datasets. The per-class results of GeneralAD on the datasets MVTec-AD [2], MVTec-LOCO [1], VisA [11], and MPDD [5].

MVTec-AD		MVTec-LOCO		VisA		MPDD	
Class	I-AUROC (%)	Class	I-AUROC (%)	Class	I-AUROC (%)	Class	I-AUROC (%)
tile	100	screw bag	74.4	candle	95.3	bracket black	93.8
bottle	100	pushpins	77.1	capsules	93.4	bracket brown	97.2
cable	98.8	juice bottle	93.7	cashew	92.9	bracket white	99.4
capsule	97.6	breakfast box	90.9	chewinggum	99.4	connector	96.7
carpet	99.8	splicing connectors	88.2	fryum	95.7	metal plate	100
grid	100			macaroni1	96.8	tubes	99.7
hazelnut	99.9			macaroni2	89.1		
leather	100			pcb1	97.2		
metal nut	100			pcb2	97.7		
pill	95.1			pcb3	96.5		
screw	96.9			pcb4	99.3		
toothbrush	100			pipe fryum	97.7		
transistor	100						
wood	99.9						
zipper	100						
All avg.	99.2		84.9		95.9		97.8

References

1. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision* **130**(4), 947–969 (2022)
2. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 9592–9600 (2019)
3. Elson, J., Douceur, J.R., Howell, J., Saul, J.: Asirra: a captcha that exploits interest-aligned manual image categorization. *CCS* **7**, 366–374 (2007)
4. Intel: Intel Image Classification. <https://www.kaggle.com/datasets/puneet6060/intel-image-classification/data> (2019)
5. Jezek, S., Jonak, M., Burget, R., Dvorak, P., Skotak, M.: Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In: *2021 13th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*. pp. 66–71. IEEE (2021)
6. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: *Proceedings of the IEEE international conference on computer vision workshops*. pp. 554–561 (2013)
7. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
8. Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. *arXiv e-prints* pp. arXiv–1306 (2013)
9. Mirzaei, H., Salehi, M., Shahabi, S., Gavves, E., Snoek, C.G.M., Sabokrou, M., Rohban, M.H.: Fake it until you make it : Towards accurate near-distribution novelty detection. In: *The Eleventh International Conference on Learning Representations* (2023)
10. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv e-prints* pp. arXiv–1708 (2017)
11. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: *European Conference on Computer Vision*. pp. 392–408. Springer (2022)