

Supplementary note for “Zero-shot versus Many-shot: Unsupervised Texture Anomaly Detection”

Toshimichi Aota
Sanoh Industrial Co., Ltd.
t.aota@sanoh.com

Lloyd Teh Tzer Tong
Sanoh Industrial Co., Ltd.
t.lloydteh@sanoh.com

Takayuki Okatani
Tohoku University / RIKEN AIP
okatani@vision.is.tohoku.ac.jp

A. Pixel-level Detection Accuracy

Table S1 shows the pixel-level detection accuracy of the proposed method, SPADE [6], and PatchCore [17]. Figure S1 shows the comparison of our method with SPADE [6], which is similar to that with PatchCore [17] in the main paper. We show these figures here due to lack of space of the main paper.

Table S1. Pixel-level detection accuracy in AUROC of our proposed zero-shot method and two non-zero-shot methods (SPADE [6] and PatchCore [17]) with different numbers of normal images. All[†] indicates the accuracy reported in the original paper.

Method	Ours	SPADE						PatchCore					
		0	1	5	10	100	All	All [†]	1	5	10	100	All
average	97.5	91.5	93.0	92.8	95.0	95.3	92.9	89.1	92.2	94.8	96.8	98.8	98.8
carpet	99.1	97.7	97.6	97.7	97.6	97.6	97.5	98.9	98.9	98.9	98.8	98.8	99.0
grid	99.1	74.4	81.6	80.8	92.2	93.1	93.7	59.6	74.7	87.6	97.5	98.1	98.7
leather	99.5	98.9	98.9	98.9	98.9	98.9	97.6	99.1	99.1	99.1	99.0	99.0	99.3
tile	93.2	91.7	91.8	91.5	91.5	91.6	87.4	95.0	95.1	95.2	95.0	95.0	95.6
wood	96.5	95.0	94.9	94.9	95.0	95.1	88.5	92.8	93.3	93.4	93.7	93.9	95.0

B. DTD-Synthetic Dataset

Table S2 shows the number of images for each textures synthesized from the DTD dataset. The images are arranged in the same manner as MVTEC AD dataset.

Table S2. DTD-Synthetic datasets

Textures	Image count		
	Train	Test (Good)	Test (Bad) / Ground truth
Blotchy_099	100	20	80
Fibrous_183	100	20	80
Marbled_078	100	20	80
Matted_069	100	20	79
Mesh_114	100	36	80
Perforated_037	100	20	80
Stratified_154	100	20	80
Woven_001	100	30	70
Woven_068	100	51	79
Woven_104	100	20	80
Woven_125	100	20	79
Woven_127	100	80	80

C. Results on DTD-Synthetic for the proposed method and SPADE

Figure S1 shows the image-level AUROC on DTD-Synthetic for the proposed method and SPADE. The results show that SPADE is unable to correctly classify the anomalous regions for most of the textures due to the reasons explained in Sec. 3. This becomes more obvious for Category-1 textures (see Sec. 5.2).

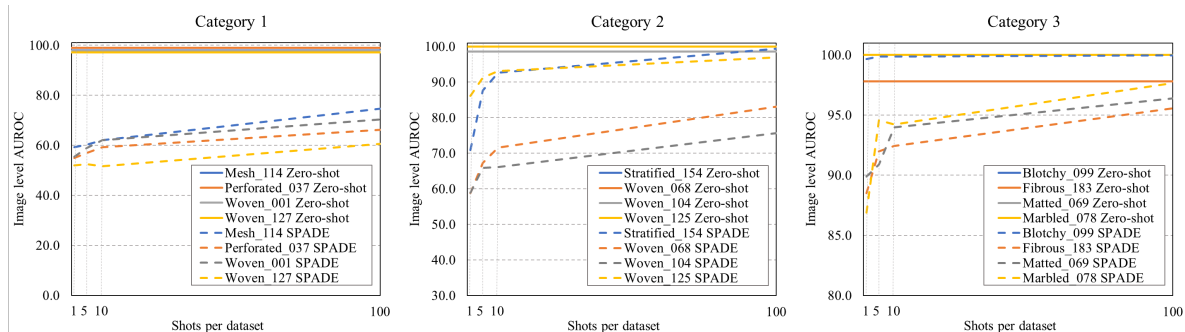


Figure S1. Detection accuracy (image-level AUROC) on DTD-Synthetic by the proposed method (solid lines) and SPADE (dotted lines) with different numbers N of normal images. As our method does not use a normal image, accuracy does not change with N . Some results from the proposed method (solid lines) are overlapping with each other as the AUROC achieved is close to 100%.

D. Dependency on the number K of nearest neighbors

Figure S2 shows the dependency of K on detection accuracy. Our method uses the average distance d_{ij} as anomaly score at image position $(i; j)$, which is the average distance from the feature vector f_{ij} at $(i; j)$ to its K nearest neighbors in the feature space. The five textures from MVTEC AD are used. It is seen that the detection accuracy is stable around from 200 to 1200, showing its insensitivity to K .

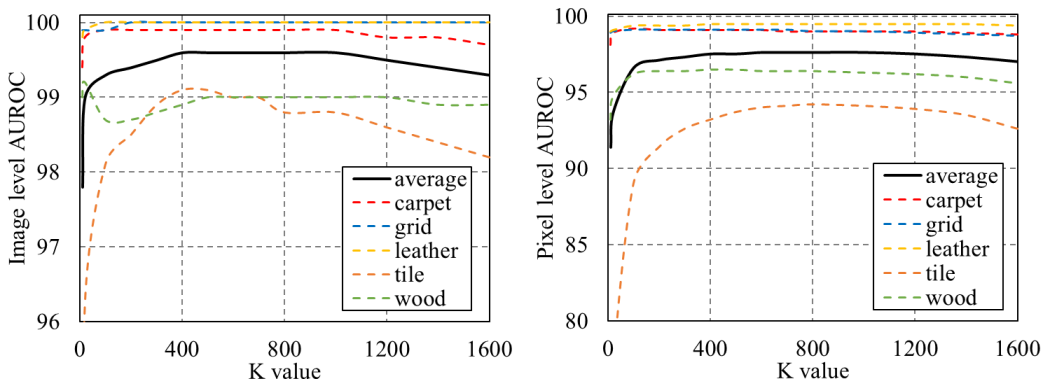


Figure S2. Dependency of detection accuracy on K . Left: Image-level AUROC. Right: Pixel-level AUROC.

E. Choice of Layer(s) for Feature Extraction

We conduct experiments to examine the dependency of detection accuracy on which layer(s) we choose for extracting feature vectors. The five textures from MVTEC AD are used. Following previous studies [17], choosing the output of the second and third blocks (out of the four blocks) of WideResnet-50-2, we evaluate our method with each of them and their combinations. Figure S3 shows the results. It is seen that choosing the second block feature is the most important; its combination with other layer features often contributes to a slight improvement in pixel-level accuracy but does not affect the results that much.

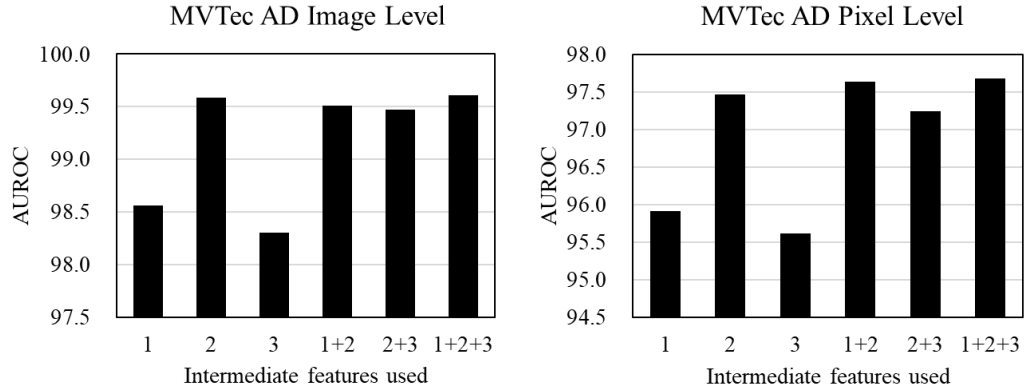


Figure S3. Results of the proposed method with different intermediate layers for feature extraction. Left: Image-level AUROC. Right: Pixel-level AUROC.

F. DTD Textures with High (I)

Figure S4 shows several examples of DTD images for which the maximum anomaly score (I) (defined in (3) of the main paper) is high. It is seen that these textures, even without anomaly, lack homogeneity and tend to yield high anomaly scores in some local regions. As explained in Sec. 4.3, we can predict our method will not work well for the images with high (I) . It does not work for these images, as shown in Fig. 5 of the main paper.

Figure S4. Examples of DTD images with high (I) and their anomaly score map. These are anomaly-free images; they are cropped at random position and orientation from the original images and are not added synthetic anomalies.