

Image Size		128		256						
Category/Method		MKD [33]	Ours	US [4]	MF [40]	SPADE [7]	PaDiM [8]	RIAD [46]	CutPaste [23]	Ours
Textures	Carpet	95.6/-	98.1/95.3	-87.9	-87.8	97.5/94.7	<b>99.1/96.2</b>	96.3/-	98.3/-	98.9/ <b>97.0</b>
	Grid	91.8/-	97.3/92.6	-95.2	-86.5	93.7/86.7	97.3/94.6	98.8/-	97.5/-	<b>99.3/97.6</b>
	Leather	98.1/-	99.0/98.6	-94.5	-95.9	97.6/97.2	99.2/97.8	99.4/-	<b>99.5/-</b>	99.4/ <b>99.1</b>
	Tile	82.8/-	92.6/84.8	<b>-94.6</b>	-88.1	87.4/75.9	94.1/86.0	89.1/-	90.5/-	<b>95.6/90.6</b>
	Wood	84.8/-	92.1/82.3	<b>-91.1</b>	-84.8	88.5/87.4	94.9/ <b>91.1</b>	85.8/-	<b>95.5/-</b>	95.3/90.9
	<i>Average</i>	<i>90.6/-</i>	<i>95.8/90.7</i>	<i>-92.7</i>	<i>-88.6</i>	<i>92.9/88.4</i>	<i>96.9/93.2</i>	<i>93.9/-</i>	<i>96.3/-</i>	<i><b>97.7/95.0</b></i>
Objects	Bottle	96.3/-	98.2/94.7	-93.1	-88.8	98.4/95.5	98.3/94.8	98.4/-	97.6/-	<b>98.7/96.6</b>
	Cable	82.4/-	97.8/90.5	-81.8	<b>-93.7</b>	97.2/90.9	96.7/88.8	84.2/-	90.0/-	<b>97.4/91.0</b>
	Capsule	95.9/-	96.5/87.2	<b>-96.8</b>	-87.9	<b>99.0/93.7</b>	98.5/93.5	92.8/-	97.4/-	98.7/95.8
	Hazelnut	94.6/-	98.8/89.2	<b>-96.5</b>	-88.6	<b>99.1/95.4</b>	98.2/92.6	96.1/-	97.3/-	98.9/95.5
	Metal Nut	86.4/-	96.6/84.1	-94.2	-86.9	<b>98.1/94.4</b>	97.2/85.6	92.5/-	93.1/-	97.3/92.3
	Pill	89.6/-	97.0/90.0	-96.1	-93.0	96.5/94.6	95.7/92.7	95.7/-	95.7/-	<b>98.2/96.4</b>
	Screw	96.0/-	98.3/94.4	-94.2	-95.4	98.9/96.0	98.5/94.4	98.8/-	96.7/-	<b>99.6/98.2</b>
	Toothbrush	96.1/-	98.2/86.7	-93.3	-87.7	97.9/93.5	98.8/93.1	98.9/-	98.1/-	<b>99.1/94.5</b>
	Transistor	76.5/-	97.6/85.2	-66.6	<b>-92.6</b>	94.1/87.4	<b>97.5/84.5</b>	87.7/-	93.0/-	92.5/78.0
	Zipper	93.9/-	97.0/92.3	-95.1	-93.6	96.5/92.6	<b>98.5/95.9</b>	97.8/-	99.3/-	98.2/95.4
	<i>Average</i>	<i>90.8/-</i>	<i>97.6/89.4</i>	<i>-90.8</i>	<i>-90.8</i>	<i>97.6/93.4</i>	<i>97.8/91.6</i>	<i>94.3/-</i>	<i>95.8/-</i>	<i><b>97.9/93.4</b></i>
<i>Total Average</i>		<i>90.7/-</i>	<i>97.0/89.9</i>	<i>-91.4</i>	<i>-90.1</i>	<i>96.5/91.7</i>	<i>97.5/92.1</i>	<i>94.2/-</i>	<i>96.0/-</i>	<i><b>97.8/93.9</b></i>

Table 2. *Anomaly Localization* results with AUROC and PRO on **MVTec** [3]. AUROC represents a pixel-wise comparison, while PRO focuses on region-based behavior. We show the best results for AUROC and PRO in bold. Remarkable, our approach is robust and represents state-of-the-art performance under both metrics.

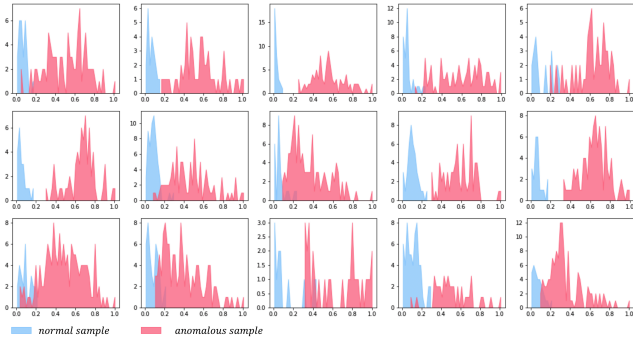


Figure 5. Histogram of anomaly scores for all categories of MVTEC [3] (x-axis: anomaly score from 0 to 1, and y-axis: count).

anomaly region.

**Complexity analysis.** Recent pre-trained model based approaches achieve promising performance by extracting features from anomaly-free samples as a measurement [7, 8]. However, storing feature models leads to large memory consumption. In comparison, our approach achieves better performance depending only on an extra CNN model. As shown in Tab. 3. Our model obtain performance gain with low time and memory complexity.

Methods	Infer. time	Memory	Performance
SPADE (WResNet50)	1.40	1400	85.5/96.5/91.7
PaDiM (WResNet50)	0.95	3800	95.5/97.5/92.1
<b>Ours (WResNet50)</b>	<b>0.31</b>	<b>352</b>	<b>98.5/97.8/93.9</b>

Table 3. Comparison of pre-trained based approaches in terms of inference time (second on Intel i7), memory usage (MB), and performance (AD-AUROC/AL-AUROC/AL-PRO) on MVTEC [3].

**Limitations.** We observe that the localization performance on the *transistor* dataset is relatively weak, despite the good AD performance. This performance drop is caused by misinterpretation between prediction and annotation. As shown in Fig. 6, our method localize the misplaced regions, while the ground truth covers both misplaced and original areas. Alleviating this problem requires associating more features with contextual relationships. We empirically find that a higher-level feature layer with a wider perceptive field can improve the performance. For instance, anomaly detection with the second and third layer features achieves 94.5% AUROC, while using only the third layer improve the performance to 97.3%. In addition, reducing image resolution to  $128 \times 128$  also achieves 97.6% AUROC. We present more cases of anomaly detection and localization, both positive and negative, in the *supplementary material*.

## 4.2. One-Class Novelty Detection

To evaluate the generality of proposed approach, we conduct *one-class novelty detection* experiment on 3 semantic datasets [32], **MNIST** [22], **FashionMNIST** [41] and **CIFAR10** [20]. MNIST is a hand-written digits dataset from numbers 0-9. FashionMNIST consists of images from 10 fashion product classes. Both datasets includes 60K samples for training and 10K samples for testing, all in resolution of  $28 \times 28$ . CIFAR10 is a challenging dataset for novelty detection because of its inclusion of diverse natural objects. It includes 50K training images and 10K test images with scale of  $32 \times 32$  in 10 categories.

Following the protocol mentioned in [27], we train the model with samples from a single class and detect novel samples. Note that the novelty score is defined as the sum