



That's BAD: blind anomaly detection by implicit local feature clustering

Jie Zhang¹ · Masanori Suganuma¹ · Takayuki Okatani^{1,2}

Received: 13 July 2023 / Revised: 13 July 2023 / Accepted: 18 January 2024 / Published online: 2 March 2024
© The Author(s) 2024

Abstract

Recent studies on visual anomaly detection (AD) of industrial objects/textures have achieved quite good performance. They consider an unsupervised setting, specifically the one-class setting, in which we assume the availability of a set of normal (i.e., anomaly-free) images for training. In this paper, we consider a more challenging scenario of unsupervised AD, in which we detect anomalies in a given set of images that might contain both normal and anomalous samples. The setting does not assume the availability of known normal data and thus is completely free from human annotation, which differs from the standard AD considered in recent studies. For clarity, we call the setting blind anomaly detection (BAD). We show that BAD can be converted into a local outlier detection problem and propose a novel method named PatchCluster that can accurately detect image- and pixel-level anomalies. Experimental results show that PatchCluster shows a promising performance without the knowledge of normal data, even comparable to the SOTA methods applied in the one-class setting needing it.

Keywords Anomaly detection · Anomaly localization · Unsupervised learning

1 Introduction

In this paper, we consider visual anomaly detection and localization of industrial objects and textures. Anomaly detection (AD) [1–3] is the task of detecting anomalous images or patterns that are out of the distribution of normal images or patterns. AD for industrial applications often requires distinguishing small differences between normal and anomalies [4–6]; see examples from a standard benchmark dataset, MVTec AD, in Fig. 1.

As anomalies can appear with countless types, and the majority are usually the normal samples in a manufacturing line, the one-class unsupervised setting has drawn the most attention from the research community. This setting assumes a set of anomaly-free images for training, which are selected by human experts, and we detect anomalies at test time.

Since the features from the standard pre-trained deep models were ‘rediscovered’ to be effective for the task [7], recent studies employing them have achieved higher and higher performances on existing public benchmarks. Those features are proved to be representative enough for local image regions, even without any adaptation to the anomaly detection datasets at hand [7, 8]. While visual unsupervised AD appears to be a solved problem due to these successes, attention has also been paid to more challenging problems, such as few-shot AD [9, 10] and developing a unified model that can detect anomalies for multiple different classes of objects/textures [11]. However, these studies continue to consider the one-class setting, assuming a perfect set of anomaly-free samples to be available, which usually needs manual annotation by manufacturing experts.

In this paper, we consider yet another scenario of unsupervised AD, which does not need any human annotations. Specifically, we consider the problem where we are not given the knowledge of normal samples for training; we want to detect anomalies in an input set of samples that might contain both normal and anomalous samples. Note that traditional machine learning often calls this setting ‘unsupervised AD’ and the above one ‘semi-supervised AD’ [12, 13], unlike recent studies in computer vision. For clarity, we call the setting blind anomaly detection (BAD) in this paper. The recent

✉ Takayuki Okatani
okatani@vision.is.tohoku.ac.jp

Jie Zhang
jzhang@vision.is.tohoku.ac.jp

Masanori Suganuma
suganuma@vision.is.tohoku.ac.jp

¹ Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan

² RIKEN Center for AIP, Tokyo, Japan

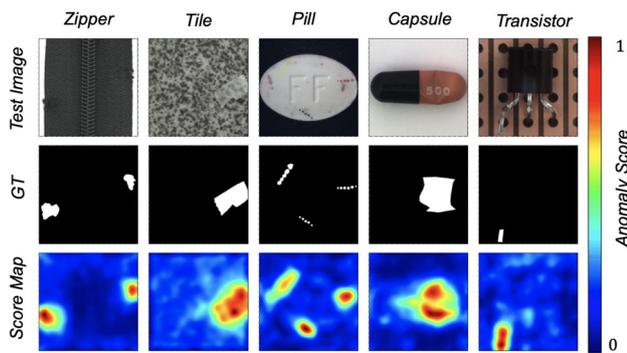


Fig. 1 Examples of industrial anomalies from the MVTec AD and the detection results of PatchCluster-25% without using the training data

unsupervised AD methods mentioned above are not designed for BAD and cannot directly be applied to the problem.

We then consider BAD a local patch outlier detection problem and introduce PatchCluster, which does not require human annotation and could automatically detect anomalies under BAD settings. We make the assumption that normal local features follow dense distributions and have small distances from each other in the feature space. Under this assumption, we propose to use local patch features to implicitly estimate the local feature distribution and use the average distance as the abnormality score. Unlike previous patch distribution modeling-based methods that assume features in the same spatial location follow the same distribution [14], we cluster local patch features for the same *contextual* location by distance-based nearest neighbor searching.

We show the effectiveness of PatchCluster and prove our assumption through comprehensive experiments and analysis. Revealing the strict restriction of spatial location-based modeling, PatchCluster is robust to spatial translation and rotation. PatchCluster achieves 95.7% and 95.9% average image-level and pixel-level anomaly scores on MVTec AD dataset without reaching any normal training data.

2 Related work

One-class anomaly detection, also known as one-class novelty detection [15, 16], is a long-standing problem in computer vision. Prior arts mainly focus on image-level outlier detection where the anomalous samples follow distributions of other semantic categories, e.g., detecting dog images for the cat category. Representation learning-based methods that could effectively learn the global contextual information are employed to tackle this problem, of which deep autoencoders (AE) [4, 17] and generative adversarial networks (GANs) [16, 18] are popular choices.

In industrial manufacturing scenarios, however, anomalies will generally occur in confined areas on a specific kind

of product, making the anomalous samples very close to the normal data distribution and the task more challenging. Recent one-class anomaly detection benchmarks [6, 19, 20] providing normal real-world industrial products for training draw lots of research attention and lots of attempts are paid to utilizing ImageNet [21] pre-trained models that could extract representative features and conduct industrial anomaly detection in a local patch feature-based manner. PaDiM [14] explicitly models the feature distribution at each spatial location. However, the assumption that patch features at the same spatial location follow the same distribution is too strict. SPADE [22] creates a feature memory bank from the normal training data and assigns anomaly scores to test image features by kNN search. The image-level anomaly scores are still based on the global image distances. PatchCore [8] propose to use locally aware features to retain more contextual information and further utilize the greedy search method to reduce the size of the memory bank. There are also many approaches based on top of the pre-trained features that try to transfer the knowledge of normal features to a student network [23–25] or estimate the distribution of normal features by flow-based methods [26–28].

While recent SOTA methods have shown nearly perfect anomaly detection performance on the public benchmark, MVTec AD dataset [19], e.g., PatchCore-Ensemble achieves a 99.6% image-level AUROC score on the MVTec AD, to the best of our knowledge, no attention has been paid to exploring achieving comparable performance without any human annotations. The one-class unsupervised anomaly detection setting requires specialists to annotate a set of normal images for training, especially for various industrial products. Under the BAD setting, most of the learning-based methods cannot be utilized directly. It should be noted that although some works have studied anomaly detection from noisy data [29, 30], i.e., anomaly-free training set contaminated by wrongly labeled anomalous samples, they are still under the one-class setting and need image labels from the human.

The methods most related to ours are PaDiM [14] and PatchCore [8]. We estimate the local patch feature distributions by kNN search, without explicitly modeling the distribution and revealing the overly strict assumptions of PaDiM. PatchCore only uses the nearest neighbor for anomaly scoring, which is sensitive enough under the one-class setting; however, we show that there may be several candidates that are too close to the anomalous feature in the memory bank and significantly decrease the anomaly detection sensitivity.

3 Blind anomaly detection (BAD)

We propose the task of blind anomaly detection (BAD), which involves identifying anomalies from a set of unknown

Table 1 Statistical overview of the proposed blind anomaly detection settings on the MVTec AD. We report the mean value over the 15 categories

	Normal images	Anomaly images	% of normal images	% of anomalous pixels
Mix	273	83	77	1.1
Test	31	83	27	3.3
Ano	0	83	0	4.7

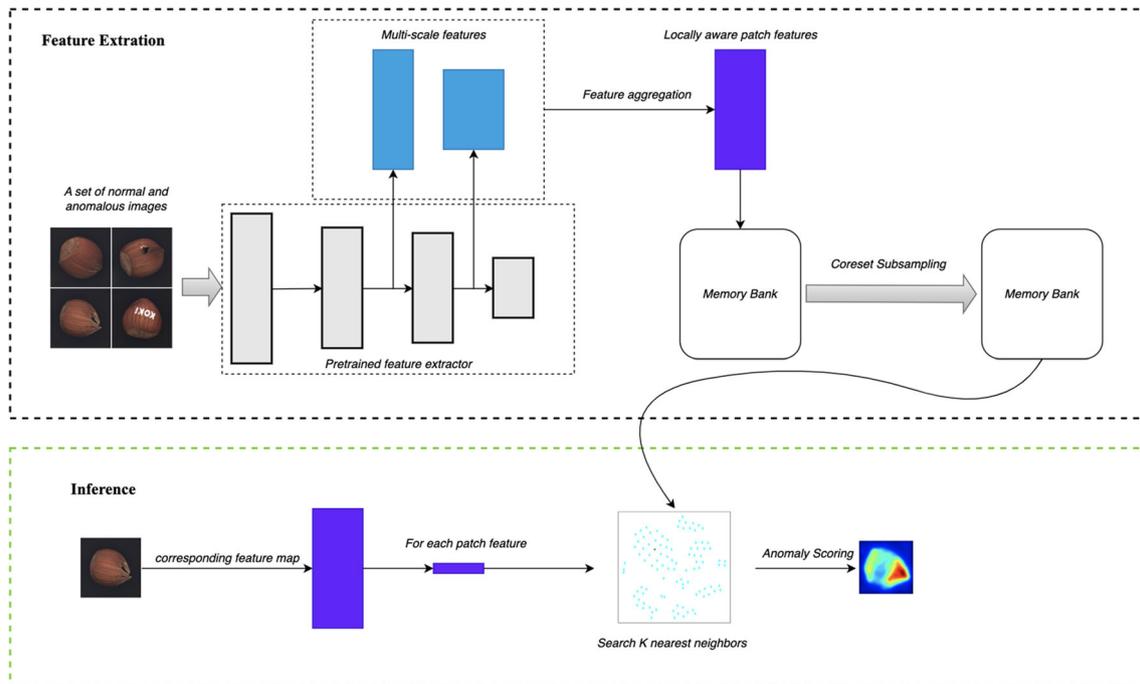


Fig. 2 Overview of the proposed PatchCluster. Given a mixed set of normal images and anomalous images, we use a pre-trained feature extractor to extract features at multiple scales. The multi-scale features for each image are then aggregated into one locally aware peach feature map and added to the memory bank. An optional coreset subsampling

process could be utilized to reduce the size of the memory bank. During inference, the K nearest neighbors for one patch feature are clustered to estimate the local patch feature distribution. Without explicitly modeling the distribution, the average distance to the K neighbors is used as the anomaly score

images that may be either normal or anomalous. Given a set of images $\mathcal{S} = I_1, I_2, \dots, I_n$, the objective is to identify anomalous images and localize the defective regions in each image. In this task, we do not use any image- and pixel-level annotations for training; we are not given any knowledge of normal or abnormal samples. Thus, recent AD methods cannot be applied directly to the BAD task.

As there is currently no comprehensive real-world dataset that aligns with this task, we formulate several variations of the BAD task using the existing unsupervised anomaly detection benchmark, MVTec AD dataset. The MVTec AD dataset [19] comprises 5 texture categories and 10 object categories and is designed for one-class anomaly detection. It includes 3629 normal images for training and 1725 normal and anomalous images for testing. We have formulated three BAD settings: MVTec AD-Mix, MVTec AD-Test, and MVTec AD-Ano. The MVTec AD-Mix setting merges the training and test images for each category, while the MVTec

AD-Test setting is the original test dataset. The MVTec AD-Ano setting is the most extreme case, in which it includes only the anomalous images from the test dataset.

We present a statistical overview of the proposed BAD settings in Table 1. Under the BAD scenarios, it is not necessary to build separate training and testing splits, as opposed to supervised learning and one-class unsupervised learning. In the Ano setting, there are no normal images. Given this, the question arises whether it is possible to not only classify the two classes of images but also determine which class is normal, even though all images are anomalous. In other words, is it possible to assign low anomaly scores to normal images, even if they account for a small portion, and assign high anomaly scores to anomalous images, even though all the images contain anomalies?

Considering the specific scene, industrial manufacturing lines, we could make use of the prior knowledge that most anomalies of industrial products will occur in subtle areas,

then for one anomaly image, it is more likely that the majority of the pixels are still normal. We also show the mean percentage of anomalous pixels over 15 categories in Table 1. The prior knowledge that the normal pixels will always account for the majority makes it possible for blind anomaly detection.

4 Local outlier detection

We then introduce the proposed PatchCluster for BAD. It is an extension of existing feature memory bank-based methods and assigns anomaly scores to each patch feature by local patch feature clustering, which makes it suitable for either one-class setting or blind anomaly detection.

4.1 Revisiting memory bank-based methods

Under the one-class setting, building a memory bank using mid-scale features extracted by a deep pre-trained model and then applying the simple nearest neighbor search for anomaly scoring could achieve nearly perfect anomaly detection performance. We also build a feature memory bank first as the extracted features have been shown to be representative enough for anomaly detection.

Given an image I_i from the dataset \mathcal{S} , the l -th layer of the pre-trained model E extracts feature map $F_i^l \in \mathbb{R}^{W^l \times H^l \times C^l}$ for the image. The feature vector $f_i^l(w, h) \in \mathbb{R}^{C^l}$ at spatial location (w, h) ($w = 1, \dots, W^l, h = 1, \dots, H^l$) is employed as the image patch representation. The pre-trained model could effectively extract features from low levels to high semantic levels. As the low-level features may be too generic and the high-level features are source-domain biased, the multi-scale features are concatenated along the channel dimension in PatchCore to get one single feature representation $(f_i(w, h))$ at each location.

To increase the effective receptive field of the pre-trained features while avoiding introducing ImageNet-biased knowledge, PatchCore further employs adaptive average pooling to fuse the feature vector with its neighboring features in the feature map.

The full feature memory bank \mathcal{M} is then created for all the images in the dataset \mathcal{S}

$$\mathcal{M} = \{f_i(w, h)\},$$

$$i \in \{1, 2, \dots, n\}, w \in \{1, \dots, W^l\}, h \in \{1, \dots, H^l\}. \quad (1)$$

To reduce the size of the memory bank, SPADE creates a memory bank for one test image using its nearest images in the dataset, while PatchCore subsamples the memory bank into a coreset by greedy subsampling.

4.2 Proposed PatchCluster method

We extend our method from the feature memory bank-based arts. We first create a memory bank \mathcal{M} following PatchCore. SPADE and PatchCore score a patch feature by the distance with its nearest neighbor feature from the memory bank. It is effective enough, as there are only normal features in \mathcal{M} . However, in BAD where the memory bank contains both normal and anomalous features especially \mathcal{M} is built on top of the dataset \mathcal{S} itself, *it may also be easy to find similar patch features for the anomalous patches.*

To address the above issue, we make the assumption that 1) features that follow the same distribution have smaller distances in the feature space and they could be used to estimate the local feature distribution in a specific *contextual* location; 2) anomalies are random events and the distribution of each type of anomaly has larger variances compared to normal features. Under the above assumption, if we interpret the creating process of memory bank \mathcal{M} as a sampling problem, it is possible to estimate the normal local distributions using the local features as they are similar to each other and have high probabilities to be sampled, while it is more difficult to estimate the anomalous distributions as they have more variants and have less probability to be sampled. We then build a local feature gallery $\mathcal{G}_i(h, w)$ by searching the corresponding K nearest neighbors. Figure 2 gives an overview of the proposed PatchCluster.

Then the anomaly detection for each patch feature is converted to a local outlier detection problem. Under the assumption, normal local distributions are in a dense feature space while anomalous local distributions are relatively sparse. We then estimate the feature abnormality by the mean distance to its neighbors without explicitly modeling the distributions

$$a_{w,h} = \frac{1}{K} \sum_{j=1}^K \text{dist}(f_i(w, h), f_j), \quad j \in \mathcal{G}_i(h, w). \quad (2)$$

With a proper K , for normal patch features, it should be easy to search for enough neighbors that follow the same dense distribution. However, for an anomalous patch feature, on the one hand, the neighbors tend to have larger distances with it. On the other hand, it is also likely to fail to find enough close neighbors from \mathcal{M} . Consequently, after scoring all the patch features in an image, we get an anomaly score map for it. As the features are down-sampled compared to the input images, we up-sample the score map by bi-linear interpolation. Following the popular choice, to remove local noises, we apply a Gaussian filter with a kernel size of 4 to get the final anomaly score map.

We could simply choose the maximum score a_i^* in a score map to account for the image-level anomaly score. We find

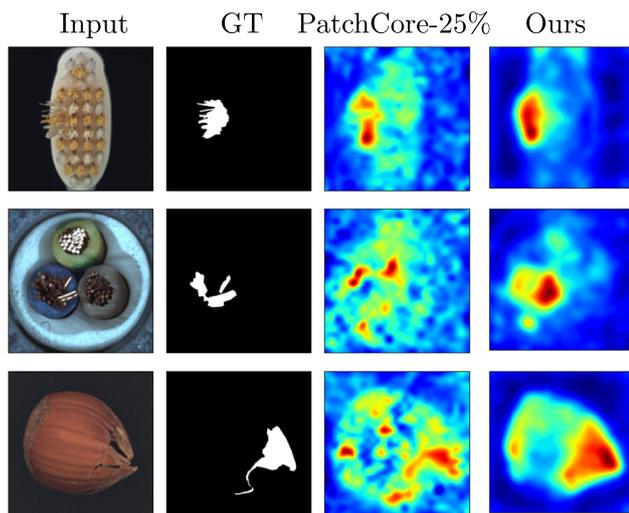


Fig. 3 Visualization examples from the toothbrush, cable, and hazelnut categories

it still robust for BAD to increase the image anomaly score if the nearest feature f^* in M of the corresponding patch feature f_i^* has a large anomaly score

$$a_i = \left(1 - \frac{\exp(a_i^*)}{\sum_{f \in \mathcal{N}(f^*)} \exp(\text{dist}(f_i^*, f))} \right) \cdot a_i^*, \quad (3)$$

where $\mathcal{N}(f^*)$ is a set of nearest neighboring patch features for f^* . The image anomaly scoring function is the same as PatchCore.

5 Experimental results and analyses

5.1 Experimental setup

5.1.1 Evaluating metrics

We use AUROC score and PRO score to evaluate the pixel-wise anomaly detection results. The two metrics are threshold-free and the PRO metric treats each anomalous region equally. We also use AUROC score to evaluate the image-level anomaly detection results except for the MVTec AD-Ano, as there is only one kind of anomalous image in this setting.

5.1.2 Implementation details

We resize the images into a 256×256 resolution and then center crop the images into 224×224 throughout our experiments. For a fair comparison, we use the same ImageNet pre-trained WideResNet-50 [31] as the feature extractor for all evaluation methods. For PaDiM [14] and PatchCore [8],

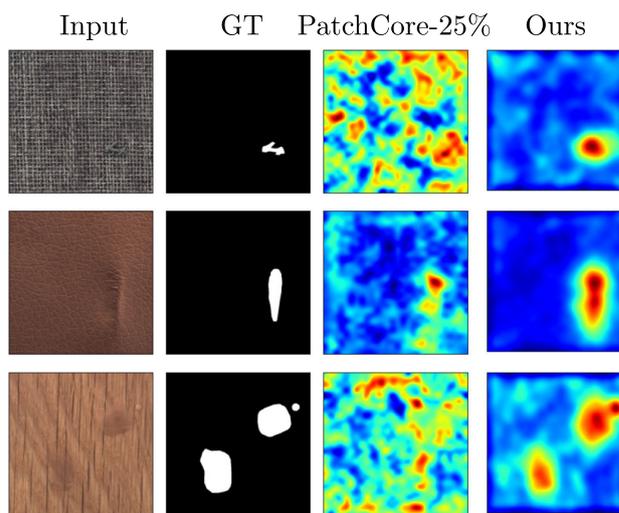


Fig. 4 Visualization examples from the carpet, leather, and wood categories

we use the same experimental settings as in their papers. We also use the intermediate features from the output of the second and third residual stages of the feature extractor, which is the same as PatchCore. We also make SPADE follow the same choice of layers to extract features, which is different from the original SPADE that also uses the first residual stage. We find this change significantly improved the inference speed and yields better performance. As SPADE first creates a memory bank for each image by image-level nearest neighbor search, we exclude the image itself when creating the memory bank to avoid including features with high similarity for both normal and anomalous features from the image itself. For PatchCore and PatchCluster especially with memory bank subsampling, however, we aim to build a unified memory bank for all test images and cannot avoid confusing features from the image itself. We could compute the anomaly score for a given patch feature using or starting from its k -th nearest neighbor. However, the optimal k for a set of images is highly dependent on the dataset size, the inner distribution of the images, and post-processing approaches on the memory bank such as coresets subsampling. To tackle this issue, we simply set k to 2 to avoid searching the feature itself as the nearest neighbor. We show that using a proper K value which has a clear choosing criterion makes PatchCluster robust to various BAD settings in 5.2 and 5.3.

We set the number of nearest neighbors K for patch feature anomaly scoring to 100 for PatchCluster-100%, which is close to the average number of test images for each category. For the coresets with subsampling ratios of 1%, 10%, and 25%, we simply reduce the K to 5, 10, and 25 without careful tuning, respectively. It should be noted that as mentioned above, we do not consider the feature itself as its neighboring feature.

Table 2 Anomaly detection and localization performance under the blind anomaly detection settings on MVTec AD

	Method	PaDiM	SPADE	PatchCore -LoF	PatchCore 10%	PatchCore -25%	PatchCore -100%	PatchCluster -25%
Mix	Image	95.2	86.8	87.8	92.7	86.6	82.9	97.5
	Pixel	97.0	92.9	88.2	84.8	85.7	87.9	98.2
	Pro	91.0	89.4	66.3	65.7	65.1	75.3	91.0
Test	Image	91.4	79.4	81.7	89.5	82.9	81	95.7
	Pixel	93.7	90.9	83.4	81	82.4	85.7	95.9
	Pro	89.5	88.5	64.4	61.8	61.9	73.7	89.5
Ano	Pixel	92.0	88.6	80.7	78.1	81.0	84.5	94.3
	Pro	88.5	87.6	63.2	59.6	61.0	72.9	88.3

Bold values indicate the best results within each experimental configuration

5.2 Blind anomaly detection on MVTec AD

We first give an overall comparison with PaDiM, SPADE, PatchCore, and the proposed PatchCluster in Table 2. The proposed PatchCluster uses the same memory bank as PatchCore. Our proposed method outperforms all competitors by a large margin under all BAD settings regarding both image-level and pixel-level anomaly detection evaluation metrics. PatchCluster-25% achieves 97.5% and 98.2% AUROC scores for anomaly detection and localization for Mix setting, which is comparable to several SOTA one-class anomaly detection approaches. Without approaching any training data under the Test setting, PatchCluster still preserves high anomaly detection performance. Under the most aggressive Ano setting, PatchCluster still achieves a 94.3% AUROC score and a PRO score of 88.3% for pixel-wise anomaly detection. As a comparison, the best image-level and pixel-level AUROC scores are only 92.7% and 92.9% under the Mix setting, at the influence of only 1.1% anomalous pixels in the dataset. It should also be noted that even for the Mix setting with the smallest portion of anomalous pixels, there are still 23% anomalous images in the dataset, which is a stringent BAD setting and practically impossible for uniform sampling from industrial product lines.

As another patch feature memory bank-based method, the modified SPADE shows better pixel-level anomaly detection than PatchCore, demonstrating the influence of confusing features from the test image itself.

PaDiM explicitly models the distributions for each fixed patch feature location. Well-modeled distributions are desired for one-class anomaly detection; however, they conflict against the performance under BAD settings as they also cover the anomalous patch features.

We also extend the PatchCore with classical Local Outlier Factor (LoF) [32] algorithm, a local-density-based anomaly scoring method. We use the full memory bank without coreset subsampling and calculate the relative local density for each patch feature according to the distances with its neighbors

as the anomaly score. We found using 100 neighbors yield better anomaly detection results, which is consistent with our assumption.

We then report the detailed detection results of PatchCluster-25% for each category under different BAD settings in Table 3. Some visualization examples under the Test setting for objects and textures are shown in Fig. 3 and Fig. 4, respectively. PatchCore fails to effectively assign high anomaly scores to anomalous patch features which leads to too many false-positive cases, i.e., normal patches could easily be detected as anomalous. PatchCluster is robust to various types of texture and object products under all BAD settings. However, it shows slightly inferior performance for categories that local spatial variations are relatively large for normal patches such as the cable, and categories with too fine-grained or large defects such as pill and transistor.

5.3 Effectiveness of local feature clustering

We first report the performance of the proposed PatchCluster-100% with different numbers of patches used for estimating the distance for the test patch feature and the local patch distribution in Table 5. PatchCluster is stable for a wide range of K . It should be noted that with $K = 1$, the PatchCluster is identical to PatchCore-100%. The proper value of K has a certain choosing criterion. From the global contextual viewpoint, each image of a certain kind of product tends to contain most kinds of local patches. Setting K near the number of total images in \mathcal{S} is likely to yield stable performance. We set K fixed to 100 for PatchCluster-100% which is approximately the average number of images for each category under the Test setting. We visualize the local feature clustering under the Test setting on two examples from one object category and one texture category in Fig. 5. The normal feature and its neighboring features tend to comprise a dense distribution, and they are close in distance to each other. However, for anomalous features, there may exist sev-

Table 3 Detailed anomaly detection results for each category under different settings

Metric	Bottle	Cable	Capsule	Carpet	Grid	Hazelnut	Leather	Metal Nut	Pill	Screw	Tile	Toothbrush	Transistor	Wood	Zipper	Avg.
Mix	Image	99.7	98.1	97.7	99.8	94.3	99.0	95.8	92.7	96.0	96.9	98.4	96.2	99.1	99.3	97.5
	Pixel	98.8	98.1	99.4	99.7	98.2	99.3	93.6	97.8	99.4	97.6	99.2	93.5	98.4	99.5	98.2
	Pro	92.9	89.9	95.9	96.8	93.1	93.2	80.8	94.1	96.9	81.3	90.2	78.3	89.1	96.7	91.2
Test	Image	98.3	97.6	92.7	99.9	91.1	99.3	91.8	93.4	84.9	99.0	96.7	96.3	97.9	96.5	95.7
	Pixel	95.9	96.0	98.6	99.2	95.3	98.0	83.8	96.5	98.1	94.7	98.3	91.0	95.4	98.7	95.9
	Pro	87.3	87.2	95.3	96.8	90.5	92.3	77.0	93.4	94.4	80.6	87.5	75.9	89.2	96.5	89.5
Ano	Pixel	94.1	93.3	98.3	98.9	94.9	96.7	80.8	96.2	97.4	92.6	97.7	81.8	94.2	98.3	94.3
	Pro	83.4	85.0	95.1	96.8	90.0	91.5	76.6	93.0	93.1	79.9	86.0	69.7	89.3	96.4	88.3

Table 4 Anomaly detection performance under the one-class setting

Method	PaDiM	SPADE	RD	PatchCore-25%	PatchCluster-100%	PatchCluster-25%	PatchCluster-10%	PatchCluster-1%
Image	95.3	85.5	98.5	99.1	97.7	98.3	98.6	97.7
Pixel	97.5	96.0	97.8	98.1	97.5	98.0	98.1	97.5
Pro	92.1	91.7	93.9	93.4	92.1	92.8	93.1	92.1

Bold values indicate the best results within each experimental configuration

Table 5 Blind anomaly detection results of PatchCluster-100% using different K values

		K	1	10	50	100	150	200
Mix	Image		82.9	92.7	97.3	97.3	97.1	96.9
	Pixel		87.9	92.9	96.7	97.2	97.4	97.5
	Pro		75.3	86.1	90.8	91.0	91.0	91.0
Test	Image		81	89.8	95.6	95.3	94.6	93.9
	Pixel		85.7	90.3	94.3	94.9	95.1	95.1
	Pro		73.7	84.1	89.5	89.5	89.3	89.0
Ano	Pixel		84.5	88.6	92.8	93.3	93.5	93.6
	Pro		72.9	83.0	88.5	88.3	87.9	87.6

Bold values indicate the best results within each experimental configuration

Table 6 Blind anomaly detection results of SPADE-Cluster using different K values

		K	1	10	20	50	100	150
Mix	Pixel		92.9	94.3	94.4	94.2	93.8	93.5
	Pro		89.4	89.4	89.0	87.9	86.5	85.5
Test	Pixel		90.9	92.4	92.4	92.1	91.8	91.4
	Pro		88.5	88.3	87.7	86.4	84.9	83.9
Ano	Pixel		88.6	89.9	89.9	90.2	90.0	89.7
	Pro		87.6	87.2	86.5	85.1	83.6	82.6

Bold values indicate the best results within each experimental configuration

Table 7 Anomaly detection results with different coresets subsampling ratios

		Ratio	1%	10%	25%	100%
Mix	Image		90.5	97.2	97.5	97.3
	Pixel		95.8	98.0	98.2	97.2
	Pro		83.9	90.3	91.2	91.0
Test	Image		76.1	94.3	95.7	95.3
	Pixel		88.8	95.7	95.9	94.9
	Pro		72.4	88.6	89.5	89.5
Ano	Pixel		85.8	93.9	94.3	93.3
	Pro		68.5	87.3	88.3	88.3

Bold values indicate the best results within each experimental configuration

eral neighboring features, but most of the neighbors tend to have large distances from them.

To further analyze the effectiveness of local feature clustering, we extend SPADE into SPADE-Cluster and experimentally verify the performance. The results are shown in Table 6. It is also obvious that with local feature clustering, the SPADE-Cluster also shows significant improvement against SPADE. The performance of SPADE-Cluster drops with K larger than 50, which is the number of nearest images used to create the memory bank for each test image.

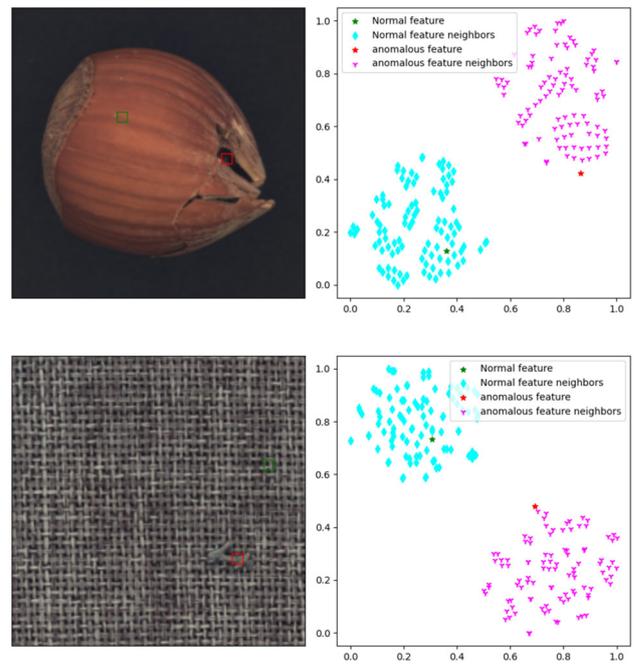


Fig. 5 Visualization of local feature clustering by nearest neighbor search

5.4 Effectiveness of memory bank subsampling

Table 7 shows the BAD performance of PatchCluster with different coresets subsampling ratios using the same greedy search method as PatchCore. We observe obvious performance drops with too small subsampling ratios, regardless of the significant memory bank size reduction and improvement of inference speed. A similar observation for PatchCore could also be found in Table 2. We argue that the local patch-based image-level anomaly scoring methods will become unreliable if the pixel-level anomaly detection ability is worse. An underlying interpretation is the difficulty of covering not only normal patch features but also anomalous patch features that are relatively sparse in the feature space with small subsampling ratios. PatchCluster-25% performs even better than PatchCluster-100% as the subsampling process, to a certain degree, plays a role of a noise feature filter.

5.5 One-class anomaly detection on MVTec AD

Table 4 shows the results of PatchCluster-25% and other methods under the one-class setting on the MVTec AD. With the patch features of the training data fully anomaly-free, the greedy-searched coresets effectively reduced the size of the memory bank while retaining strong anomaly detection ability. However, as there are no anomalous patch features in the memory bank and coresets, using a certain amount of neighboring patch features for anomaly scoring reduces the

detection sensitivity, leading to slightly inferior performance compared to PatchCore.

6 Conclusion

We introduce the blind anomaly detection (BAD) problem for industrial inspection, a task of finding fine-grained local anomalies in a set of mixed normal and anomalous images without using any human annotations. We formulate three BAD settings, Mix, Test, and Ano based on the existing industrial anomaly detection benchmark MVTEC AD dataset.

Based on the memory bank-based method, we convert BAD as a local outlier detection problem and propose PatchCluster, a method of measuring the local patch feature's distribution using the corresponding nearest neighbors in the memory bank. The proposed approach could effectively estimate the normal feature distribution, while it fails for anomalies. PatchCluster is robust to various kinds of BAD settings and shows comparable performance with current one-class anomaly detection SOTAs.

Acknowledgements This work was partly supported by JSPS KAKENHI Grant Number 20H05952 and 23H00482.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. *ACM Comput. Surv.* **41**(3), 1–58 (2009)
- Chalopathy, R., Chawla, S.: Deep learning for anomaly detection: a survey. *arXiv preprint arXiv:1901.03407* (2019)
- Xia, Y., Cao, X., Wen, F., Hua, G., Sun, J.: Learning discriminative reconstructions for unsupervised outlier removal. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1511–1519 (2015)
- Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011* (2018)
- Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but different: Semi-supervised defect detection with normalizing flows. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1907–1916 (2021)
- Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: Beyond dents and scratches: logical constraints in unsupervised anomaly detection and localization. *Int. J. Comput. Vision* **130**(4), 947–969 (2022)
- Reiss, T., Cohen, N., Bergman, L., Hoshen, Y.: Panda: Adapting pretrained features for anomaly detection and segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2806–2814 (2021)
- Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328 (2022)
- Sheynin, S., Benaim, S., Wolf, L.: A hierarchical transformation-discriminating generative model for few shot anomaly detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8495–8504 (2021)
- Huang, C., Guan, H., Jiang, A., Zhang, Y., Spratling, M., Wang, Y.-F.: Registration based few-shot anomaly detection. In: *Proceedings of European Conference on Computer Vision*, pp. 303–319 (2022). Springer
- You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. *arXiv preprint arXiv:2206.03687* (2022)
- Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., Cho, D., Chen, H.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: *Proceedings of International Conference on Learning Representations* (2018)
- Akçay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: *Proceedings of Asian Conference on Computer Vision*, pp. 622–637 (2019)
- Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: *Proceedings of International Conference on Pattern Recognition*, pp. 475–489 (2021). Springer
- Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: *Proceedings of International Conference on Machine Learning*, pp. 4393–4402 (2018). PMLR
- Perera, P., Nallapati, R., Xiang, B.: Ogan: One-class novelty detection using gans with constrained latent representations. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2898–2906 (2019)
- Pidhorskyi, S., Almoheisen, R., Doretto, G.: Generative probabilistic novelty detection with adversarial autoencoders. *Adv. Neural Inf. Process. Syst.* **31** (2018)
- Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3379–3388 (2018)
- 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)
- Bergmann, P., Jin, X., Sattlegger, D., Steger, C.: The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045* (2021)
- Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)

22. Cohen, N., Hoshen, Y.: Sub-image anomaly detection with deep pyramid correspondences. arXiv preprint [arXiv:2005.02357](https://arxiv.org/abs/2005.02357) (2020)
23. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4183–4192 (2020)
24. Bergmann, P., Sattlegger, D.: Anomaly detection in 3d point clouds using deep geometric descriptors. arXiv preprint [arXiv:2202.11660](https://arxiv.org/abs/2202.11660) (2022)
25. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9737–9746 (2022)
26. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2D normalizing flows. arXiv preprint [arXiv:2111.07677](https://arxiv.org/abs/2111.07677) (2021)
27. Rudolph, M., Wehrbein, T., Rosenhahn, B., Wandt, B.: Fully convolutional cross-scale-flows for image-based defect detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1088–1097 (2022)
28. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 98–107 (2022)
29. Jiang, X., Liu, J., Wang, J., Nie, Q., Wu, K., Liu, Y., Wang, C., Zheng, F.: Softpatch: unsupervised anomaly detection with noisy data. *Adv. Neural. Inf. Process. Syst.* **35**, 15433–15445 (2022)
30. Chen, Y., Tian, Y., Pang, G., Carneiro, G.: Deep one-class classification via interpolated gaussian descriptor. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 383–392 (2022)
31. Zagoruyko, S., Komodakis, N.: Wide residual networks. arXiv preprint [arXiv:1605.07146](https://arxiv.org/abs/1605.07146) (2016)
32. Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93–104 (2000)

Jie Zhang obtained his Bachelor's degree from Shandong University in 2014 and his Master of Science degree from Tsinghua University in 2017. He earned his Ph.D. in System Information Sciences from Tohoku University in 2023. He is currently an engineer at BYD Auto Industry Company Limited.

Masanori Suganuma received the Ph.D. degree from Graduate School of Environment and Information Sciences, Yokohama National University, in 2017. He is currently an Assistant Professor at Tohoku University. His research interests are in the field of computer vision and machine learning.

Takayuki Okatani earned his B.Eng., M.Sc., and Ph.D. degrees in Mathematical Engineering and Information Physics from the Graduate School of Engineering at the University of Tokyo in 1994, 1996, and 1999, respectively. He currently serves as a Professor in the area of computer vision at Tohoku University. In addition, he heads the Infrastructure Management Robotics Team at the RIKEN Center for Advanced Intelligence Project. With over 100 publications in peer-reviewed journals and conference proceedings, his work encompasses computer vision, deep learning, and multi-modal AI. He is an active member of several professional societies, including the IEEE Computer Society, the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, and the Society of Instrument and Control Engineers.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.