Self-Supervised Augmented Patches Segmentation for Anomaly Detection*

Jun Long[1], Yuxi Yang[2], Liujie Hua[3], and Yiqi Ou[4]

Big Data Research Institute of Central South University, Changsha City, Hunan Province, China
{junlong, yuxiyang}@csu.edu.cn

Abstract. In this paper, our goal is to detect unknown defects in high-resolution images in the absence of anomalous data. Anomaly detection is usually performed at image-level or pixel-level. Considering that pixel-level anomaly classification achieves better representation learning in a finer-grained manner, we regard data augmentation transforms as a self-supervised segmentation task from which to extract the critical and representative information from images. Due to the unpredictability of anomalies in real scenarios, we propose a novel abnormal sample simulation strategy which augmented patches are randomly pasted to original image to create a generalized anomalous pattern. Following the framework of self-supervised, segmenting augmented patches is used as a proxy task in the training phase to extract representation separating normal and abnormal patterns, thus constructing a one-class classifier with a robust decision boundary. During the inference phase, the classifier is used to perform anomaly detection on the test data, while directly determining regions of unknown defects in an end-to-end manner. Our experimental results on MVTec AD dataset and BTAD dataset demonstrate the proposed SSAPS outperforms any other self-supervised based methods in anomaly detection. Code is available at https://github.com/BadSeedX/SSAPS.

Keywords: Anomaly detection · Self-supervised learning · Data augmentation.

1 Introduction

Anomaly detection is a technique for outlier detection, commonly used for risk control[1] or workpiece surface defect detection[2]. It is a critical and long-standing problem for the financial and manufacturing industries[3]. Suffering from anomalous data scarcity and unexpected patterns, the modeling process of supervised learning is severely constrained. Therefore, unsupervised methods based on modeling of anomaly-free data are more general in industrial scenarios, and anomaly detection is usually expressed as a one-class classification

* Supported by Network Resources Management and Trust Evaluation Key Laboratory of Hunan Province, Central South University.

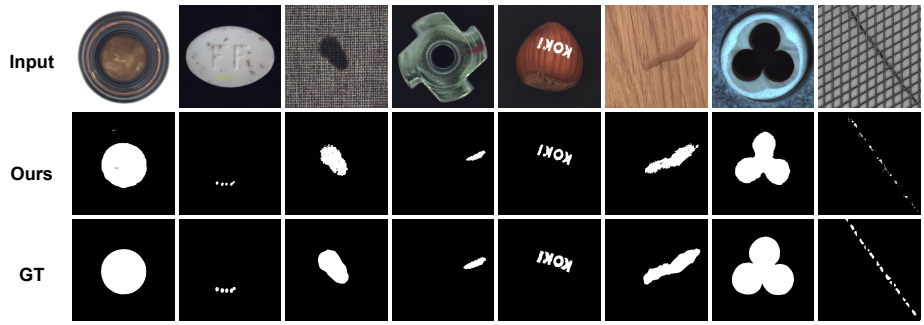

Fig. 1. Defect localization on some catagories of MVTec dataset.From top to bottom, input abnormal images, segmentation results generated by our method, and ground truth images.

problem[4,5,6], which attempts to construct the margin of normal samples. However, these approaches require complex feature matching during defect inference phase, resulting in excessive computational cost of the models. Furthermore, degradation occurs as a consequence of weak inter-class variation, also known as "hypersphere collapse"[5].

By designing a proxy task, self-supervised models[4,7,8] bypass the need for labeled data and serve as the effective proxy for supervised learning. These models preserve intra-class commonalities and extend inter-class variation by simulating anomalous patterns to learn visual representations. Nevertheless, the challenge raised by self-supervised methods is that relying on a pre-trained model risks feature extraction inconsistencies between the training source and target domains, which is insufficient to construct robust anomalous pattern margins. Due to the presence of tiny anomalous regions, models have a tendency to overfit and misclassify normal samples as abnormal samples[9]. Various data augmentation strategies[7,8,10] demonstrated a good ability to simulate anomalies in recent work. However, these approaches fail to overcome the limitation of the binary classification task for defect localization. Furthermore, these artificially created anomaly patterns are challenging to apply in real-world scenarios because of their sophisticated designs.

Given the shortcomings of above approaches, we propose an end-to-end self-supervised network based on segmentation of augmented patches. Our innovation lies in the design of a novel proxy task and the construction of a generative one-class classifier based on learned self-supervised representation to distinguish abnormal image from normal ones. Specifically, we use patches segmentation as a self-supervised proxy task to infer simulated anomalous regions by predicting augmented patches. Meanwhile, due to the diversity of real anomalies, the samples transformed by augmentation might only be considered as a rough approximation of the real defects. In fact, instead of introducing a method focused on simulating actual anomalies as is used in NSA[7], we design a generalized augmentation strategy that significantly increases irregularity of augmented patches

in order to create more complex local irregular patterns with the aim of shaping a scalable normal pattern boundary by providing a common anomlous pattern that assists the network in learning outliers in a fine-grained manner.

In this work, we tackle the problem of one-class defect detection in high-resolution images using a proxy one-class classifier constructed during the inference phase based on visual representations obtained by segmenting augmented patches. We evaluate SSAPS on the MVTec AD[2] and BeanTech AD[11] dataset, both of which derive from real-life scenarios in industrial production. With 98.1% and 96.2% AUC on each dataset separately for image-level anomaly detection, our method outperforms existing self-supervised methods. SSAPS also exhibits strong anomaly segmentation abilities. We conduct an extensive study with various proxy tasks to prove the effectiveness of prediction augmented patches for unknown defect detection. The results also suggest that the introduction of a segmentation task to extract visual representations provides a well generalization ability for anomaly detection. The following are our contributions:

• We propose a novel data augmentation strategy which creates augmented samples for self-supervised learning by replacing irregular patches in images at random to simulate generalized anomalous patterns.

• We propose an augmented patches segmentation based proxy task to train a self-supervised residual deconvolution network to find the tiny variance between normal and abnormal patterns at pixel-level with the aim of practical needs of one-class anomaly detection.

• Through extensive experimental validation, SSAPS as a simulation of semantic segmentation in vision has high accuracy in surface defect detection while better meeting the requirements of industrial scenarios in a finer-grained manner.

2 Related Work

2.1 Embedding-based approach

Embeddding-based approaches[5,12,13,14,15,16] usually use a deep network pretrained on ImageNet[17] to extract visual representations in images, and define anomaly scores for test samples through various embedding-similarity metrics. Some works use one-class classification[5], multivariate Gaussian distribution[16] or improved k-nearest neighbour[15] to evaluate images that deviate from standard representation descriptions which are considered abnormal. Traditional embeddding based models[5] consider extracting features from the entire image and work with image-level embeddings, result in failing to explain the location of the defect. Recently some work extent embeddings to patch-level to obtain a more detailed visual representation. SPADE[15], based on KNN method, detects anomalies through the pixel-level correspondence between the normal images and the test images.

2.2 Reconstruction-based approach

Reconstruction-based approaches use auto-encoder[4,18,19] or generative adversarial network[10,20,21,22,23] to train a network to reconstruct the input im-

ages. They usually focus on the difference between the reconstruction output and the original input, where images containing defects tend to produce larger reconstruction errors. OCR-GAN[22], based on an omni-frequency reconstruction convolutions framework, hires a frequency decoupling module and a channel selection module to find differences in the frequency distribution of normal and abnormal images and complete the restruction process. InTra[23], based on a deep inpainting transformer, inpaints the masked patches in the large sequence of image patches, thereby integrating information across large regions of the input image and completing a image repair task. However, the downside of these methods is that it is hard to restore details of normal patterns, due to the strong learning representation ability and the huge capacity of neural network[24], reconstruction-based approaches are still affected by the uncertainty of the anomalous regions during inference which anomalous regions in the image are unexpectedly reconstructed well.

2.3 Self-supervised learning-based approach

Self-supervised learning automatically generates supervisory signals under proxy tasks by mining the intrinsic properties of unlabelled data, thereby training the network to learn useful features that can be transfered to downstream tasks. Existing self-supervised anomaly detection methods, such as rotation prediction[25], geometric transformations prediction[26], block relative position prediction[4,27], and data augmentation classification prediction[7,8,28], share a common denominator that a classification task is preset based solely on the normal data, and the representation learned by network used for discovering unknown defects.

3 Methods

This section outlines the overall framework of our method. An overview of SSAPS is shown in Fig. 2. Following the general paradigm of self-supervised learning, SSAPS consists of a two-stage defect detection framework, aims at exploring local irregular patterns from the constructed augmented samples and attempts to segment the replaced patches, thus learning visual representations to construct a one-class anomaly classifier for inference of unknown defects in the test images. It contains several essential parts: AP module for augmenting samples by replacing irregular image patches in the original image(Sec. 3.1), self-supervised task based on patches segmentation and its loss function(Sec. 3.2), and inference phase(Sec. 3.3).

3.1 Add augmented patches to image

This subsection will concentrate on the process of constructing an augmented sample. It seems hard to move irregular image patches directly, so we save different areas of the original input image I with the help of mask image M and

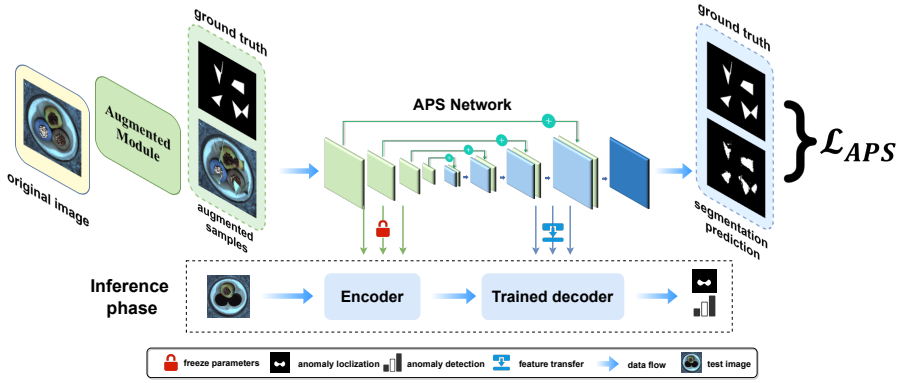

Fig. 2. An overview of our method. SSAPS is based on the framework for self-supervised learning. We get the augmented image and the ground truth image through Augmented module. The augmented image is then fed into the SSAPS network to obtain segmentation prediction. The process of the inference phase is similar to the training phase.

bottom image B, respectively, and then stack them up to achieve a partial irregular misalignment of the image. As shown in Fig. 3, the construction of augmented samples is mainly divided into the following steps.

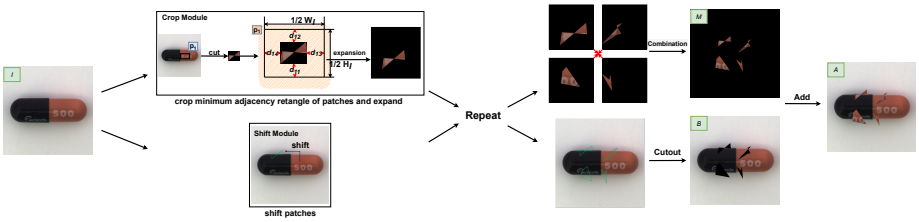

Fig. 3. Steps for the the construction of simulation anomalous samples. In the first step, mask image M consists of multiple expansions of the minimum adjacency rectangle containing irregularly augmented patches. In the second step, bottom image B is made by masking in the corresponding position of the original image. Lastly, overlay mask image and bottom image.

For M, it comes from the expansion and combination of the augmented patches. We randomly select a series of coordinate points within a specified size region of I and connect them in turn to form an irregular polygon p. Then we cut out p by a minimum adjacent rectangle and randomly extend it around a distance d with RGB(0,0,0). Through a pair of "cut-expansion" operations, we get a quarter mask with half the length and width of the original image, which

preserves a patch of I. The crop module is performed repeatedly and the quarter masks are combined into a mask image.

For B, we move p according to d and the quarter mask, ensuring that the relative position of p in B is the same as in M. In contrast to M, the image information on the outside of p is retained and all pixels on the inside of p are replaced using zero values.

For A, we overlay M and B, which resemble a pair of complementary images. The augmented image A obtained after their combination implements a random shift of the irregular patches in I. Compared with the traditional cutpaste, our method has more drastic edge dithering and provides greater convenience for the network to recognize anomaly patterns. As for the construction of the label identifying anomalous region, we imitate the ground-truth image in the test set with a binary value representing the category of each pixel, and create a matrix G of the same size as I, all points inside p are found in G to modify the category at the corresponding pixel location.

3.2 Augmented patches segmentation

In the traditional convolution architecture, the input image is converted into a multi-channel feature map after multiple downsampling. The pooling layer and the linear layer work collaboratively to convert it into a one-dimensional vector and provide it to the fully connected layer for downstream tasks. However, the flattened feature map fails to retain the anomaly regions information extracted by the convolution network, making defect localization difficult.

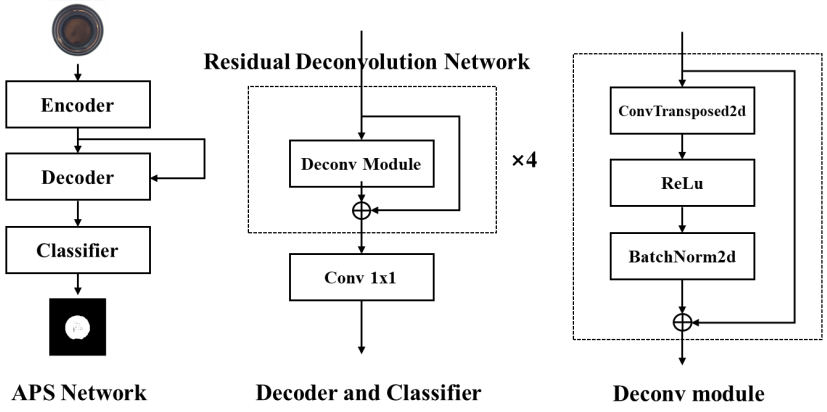

Fig. 4. The detailed structure of Augmented Patches Segmentation Network. Original image is extracted by the encoder(ConvNeXt) to obtain feature maps, and after 4 stacked inverse residual convolution modules to obtain a segmentation result of the same size as the input. Finally, the segmentation prediction is output by the classifier.

In our method, we make full use of the receptive field information from each stage of the convolution. Augmented Patches Segmentation network consists of a symmetrical encoder block and a decoder block. As shown in Fig. 4, we use ConvNeXt[29] as the encoder and save the feature map generated at each stage. Residual Deconvolution Network plays the role of decoder, which is stacked by several deconvolution modules, each module contains a BN(Batch Normalization) layer and a ReLU activation function. The deconvolution operation doubles the length and width of the feature map produced by the corresponding convolutional layers, while halving the channel. The output of each deconvolution module is then added with the upper stage's convolutional feature map as the input to the next layer. We treat the final output of the RDN as the extracted visual features, and perform anomaly classification on each pixel by a simple convolutional layer to obtain a binary image as the segmentation prediction of augmented patches.

$$\mathcal{L}_{APS} = E_{x \in X}\{BCE[seg(I), 0] + BCE[seg(A), (0, 1)]\}. \tag{1}$$

The entire self-supervised network is optimized by SGD(Stochastic Gradient Descent), and the training data is obtained from Sec. 3.1, consisting of normal image and augmented image, with corresponding binary image labelled with the augmented regions. We use Eq. 1 as objective functions for self-supervised learning and compare the segmentation prediction with the corresponding ground truth, where X is the set of normal samples, $seg(\cdot)$ is the Augmented Patches Segmentation network, $E(\cdot)$ refers to the expectation and $BCE(\cdot, \cdot)$ refers to a binary cross-entropy loss. This loss enables model learn the representation of normal patterns whereas being sensitive to outliers.

This completes the "augmented or non-augmented" segmentation prediction for all pixels in the training image, which is a simulation of classifying "normal or abnormal" pixels in the test image. As demonstrated in Sec. 4.2, we believe that the completion of this siamese task is vital for anomaly detection and segmentation.

3.3 Inference

In this section, we will go over in detail how our proposed method detects and infers unknown defects. As demonstrated above, we use a holistic approach to generate artificially simulated abormal samples to provide supervised signals for the proxy task, and then learn visual representations by segmenting augmented regions to build a self-supervised feature extractor f. We introduce the Mahalanobis distance[30] to measure how many standard deviations away a given sample is from the mean of a distribution of normal samples[31], and use it as the anomaly score to detect real unknown defects in the test image.

We define $f(*)$ as the visual representation mapping obtained after the image has been processed by the self-supervised feature extractor. Eq. 2 and Eq. 3 are used to compute the mean of distribution X_m and the covariance X_{conv} for the normal samples, where X is the set of normal samples with the length of N,

$E(*)$ refers to the expectation.

$$X_m = E_{x \in X}(f(x)). \tag{2}$$

$$X_{conv} = \frac{1}{N-1} \left(f(x) - X_m\right) \cdot \left(f(x) - X_m\right)^\top. \tag{3}$$

At this point, we can obtain a feature hypersphere similar to the one obtained in the one-class classfication methods as a potential true pattern of train images. Then we calculate the anomaly score for the test image θ via Eq. 4, by comparing the mahalanobis distance between $f(\theta)$ and X_m.

$$score = \sqrt{\left(f(\theta) - X_m\right)^\top \cdot X_{conv}^{-1} \cdot \left(f(\theta) - X_m\right)}. \tag{4}$$

4 Experiment

In order to extensively evaluate the performance of SSAPS in anomaly detection, we conducted a series of ablation experiments on unsupervised datasets MVTec AD and BeanTech AD to evaluate the contribution of individual components of the proposed method and the effectiveness of segmentation-based self-supervised tasks. Additionally, performance of SSAPS on anomaly detection is compared with other recent models based on self-supervised learning, our method exhibit superior performance and outperform any of them.

4.1 Experiment details

Our method follows the framework of the self-supervised learning[32], using ConvNeXt[29] as a feature extractor to encode the input image. We resize images in both the MVTec AD and BTAD datasets to 256×256, with each category trained for 300 iterations and the batch size set to 16, containing 8 normal samples and 8 abnormal samples. Their labels are binary images, with 0 for normal and 1 for abnormal. For the object categories in the dataset, we paste the irregular image patches closely to the centre of the image, while for the texture categories we pasted the patches evenly around because of the similarity of the entire image. More Experiment details are shown in Appendix.

4.2 Effectiveness of segmentation task

We utilize t-SNE[33] to evaluate the difference in distribution between the augmented samples and the real anomalous samples as well as the normal samples in order to better showcase the significance of the augmented patches in the segmentation task. t-SNE receives the output from the encoder and visualises representation by category. As shown in the Fig. 5, the three colors represent augmented samples, abnormal samples, and normal samples, respectively. The representation distributions of the augmented samples and the true abnormal samples overlap in space, while the augmented samples are separated from the

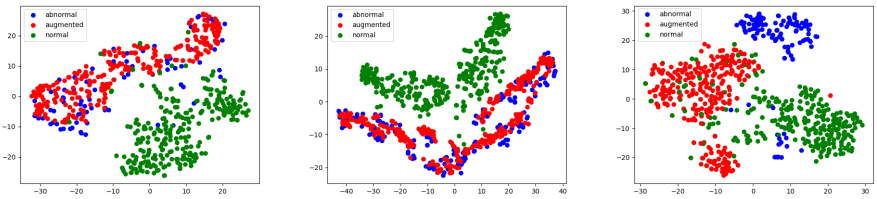

Fig. 5. t-SNE visualization of representations normal samples(green), real abnormal samples(blue), augmented samples constructed by our method(red) on some catagories in MVTEC dataset.

normal data, which not only validates our proposed abnormal sample construction method, but also confirms the effectiveness of the segmentation task. Notably, there are equally significant differences between the augmented samples and the true anomalous samples in some categories(Fig. 5.c), but this has no effect on the anomaly detection results. Predicting the transformed pixels after augmentation is, in fact, a finer-grained binary classification task in which the model focuses on inter-class differences rather than requiring that the simulated anomalous samples have the same pattern distribution as the true anomalous samples. The inclusion of anomalous simulation samples in our proposed method allows the model to explicitly learn the difference between normal and abnormal patterns rather than constructing a supervised task that tends to overfit, thus ensuring excellent generalization of our method.

4.3 Comparison with other SSL models

We believe our proposed method based on segmentation of augmented patches outperforms other existing self-supervised proxy tasks in extracting difference between normal and abnormal patterns of images. To this end, we apply various proxy tasks for image-level anomaly detection on MVTec AD and BTAD dataset, and the results are shown in the Table 1. Our approach outperforms self-supervised methods that utilize other proxy training tasks. For the MVTec AD dataset, we achieve a near-perfect result in the texture category, as well as significant improvements in the majority of categories on objects comparing to other approaches. However, one of the poorly detected object categories is pill, where the analysis of the data suggests that some of the pill surface defects in the test set are small and barely different from the normal data, resulting in difficulties in being captured by the network and causing misclassification, and our model encounters some obstacles in dealing with these defects with weak anomalous features, as is also the case to some extent with the screw category, as is shown in Fig. 6. For the BeanTech AD dataset, the AUC detection score of SSAPS is still better than all the other methods, which demonstrates the strong generality of our method.

Table 1. Comparison of our method with others for the image-level anomaly detection performance on MVTec AD and BeanTech AD dataset. The results are reported with AUROC%

MVTec AD						
Category		**R-Pred[26]**	**P-SVDD[4]**	**NSA[7]**	**CutPaste[8]**	**Ours**
object	**bottle**	95.0	98.6	97.5	98.2	**100.0**
	cable	85.3	90.3	90.2	81.2	96.9
	capsule	71.8	76.7	92.8	**98.2**	97.8
	hazelnut	83.6	92.0	89.3	98.3	**100.0**
	metal_nut	72.7	94.0	94.6	**99.9**	96.3
	pill	79.2	86.1	94.3	**94.9**	90.6
	screw	35.8	81.3	90.1	88.7	95.4
	toothbrush	99.1	100.0	99.6	99.4	**100.0**
	transistor	88.9	91.5	92.8	96.1	**97.9**
	zipper	74.3	97.9	99.5	**99.9**	98.6
object AVG		78.6	90.8	94.1	95.5	**97.4**
texture	**carpet**	29.7	92.9	90.9	93.9	**97.8**
	grid	60.5	94.6	98.5	100.0	**100.0**
	leather	55.2	90.9	100.0	100.0	**100.0**
	tile	70.1	97.8	100.0	94.6	**100.0**
	wood	95.8	96.5	97.8	99.1	**100.0**
texture AVG		62.3	94.5	97.5	97.5	**99.6**
Overall AVG		73.1	92.1	95.2	96.1	**98.1**

BeanTech AD					
Category	**R-Pred[26]**	**P-SVDD[4]**	**MemSeg[24]**	**CutPaste[8]**	**Ours**
01	80.3	95.7	98.7	**96.7**	98.3
02	50.5	72.1	87	77.4	**90.2**
03	72.6	82.1	99.4	99.3	**100**
Overall AVG	67.8	83.3	95.0	91.1	**96.2**

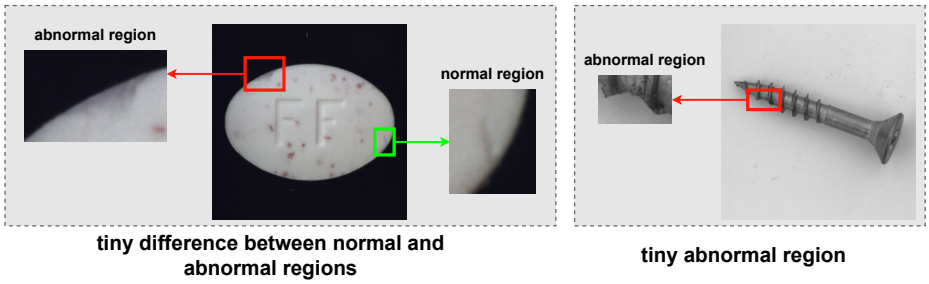

Fig. 6. Some failed detection examples of pill and screw. The red box on the left shows the small defective areas and the green box on the right shows a small part of the normal image. Surface defects are not only tiny but also differ minimally from normal regions.

4.4 Ablation Study

We conduct various further studies to provide an explainable reason what contributes to discover anomalous regions in SSAPS. We begin with performing diverse patch replacement on the normal image to explore segmentation tasks of varying complexity. The impact of various data augmentation transforms based on cutX on anomaly detection and anomaly segmentation is then compared.

Hyperparameters We evaluate the impact of the various augmented patch construction parameters, including the size, irregularity, and number of augmented patches, as well as the distribution of the pasted position, which determine the extent to the original image transformation. Table 2 indicates the detection results for various parameter on the object and texture categories, respectively. All experiments are performed under other optimal parameter conditions and the values in bold indicate the optimal results for each category of parameters. Fig. 7 and show the examples of the various transformations on some categories in the MVTec dataset.

The results in Table 2(top left) show that larger augmented patches tend to be captured. Moreover, sufficient information about the original image must be maintained in the constructed anomalous samples, due to the need to ensure a balance of positive and negative samples to avoid much more enhanced pixels than normal ones in order to allow the network to extract information about the anomalous differences from them to construct robust classification margins.

The results in Table 2(bottom left) show that a moderate increase in complexity of the pasted patches improves anomaly detection. We analyse that high complexity augmented patches increase the local irregularity of the constructed anomalous samples and make them easier to be expressed in the receptive field; however, when the complexity is excessively high (shown in the Fig. 7), the model overfits and fails to learn the weak anomalous patterns. Furthermore, for the object class, we need to paste the augmented patches in the center of

(a) Different size and number of the augmented patches

(b) Different irregularity and distribution of the augmented patches

Fig. 7. Visualization of different parameters of augmented patches on some categories in MVTEC dataset.

the image, these images consist of both object and background areas, the object usually locates in the center of the image and sometimes occupies a small proportion(e.g. capsule, screw, pill), while the background contains little anomalous valid information, when the patches are pasted into the background area (as shown in the Fig. 7), they interfere with the extraction of normal patterns in the form of noise, resulting in poor detection. For the texture category, due to the weak semantic differences in the entire image and the fact that defects can occur anywhere, the distribution of the augmented patches should be as dispersed and random as possible, and this conjecture is confirmed experimentally by the results of segmenting augmented patches randomly distributed over the entire image, which give significantly better results for defect detection than segmenting patches distributed in the center of the image.

CutX Transformations We explore other variants based on the CutX method, including cutout, cutmix and cutpaste. Cutout masks an area of the image, CutPaste masks and fills it with content from other areas of the same image while CutMix fills it with other image blocks, our method is described in Sec. 3.1. More details about cutX methods can be found in Appendix.

As the results shown in Table 3, our method demonstrates the best detection performance. On the one hand, pasting irregular patches of images increases the variability between the abnormal and normal patterns, making it easier for the network to extract features. On the other hand, capturing information from the same image allows the constructed anomalous patterns to be generalised, which prevents the network from learning the constructed anomalies as a new category and thus ignoring the real abnormal patterns.

4.5 anomaly segmentation

Our method demonstrates excellent anomaly detection performance at image-level through experimental validation. Furthermore, we experiment with defect localization in an end to end manner. Specifically, we put the test images into

Table 2. Results of different irregularity and distribution of augmented patch. The values in the patch irregularity means the range of the number of vertices of irregular polygon, the patch distribution means the degree of proximity to the centre of the image. The results are reported with AUROCs.

Category	Size of patch				Number of patch			
	0.1	**0.2**	**0.3**	**0.45**	**1**	**2**	**3**	**4**
object	96.3	96.7	**97.4**	95.9	96.5	97.1	96.8	**97.3**
texture	98.1	99.2	**99.6**	98.9	97.9	98.5	98.6	**99.4**
Overall AVG	96.8	97.5	**98.1**	96.9	96.9	97.6	97.4	**98**
Category	**Irregularity of patch**				**Distribution of patch**			
	3∼5	**5∼12**	**12∼24**	**3∼24**	**0.01∼0.5**	**0.5∼0.75**	**0.75∼0.99**	**0.01∼0.99**
object	95.9	**97.4**	97	96.8	94.3	95.9	**97.3**	95.6
texture	98.3	**99.3**	98.9	99.1	98.7	98.6	98.3	**99.3**
Overall AVG	97.6	98	97.6	97.5	95.8	96.8	97.6	**96.8**

the APS network and compare the anomaly scores of the corresponding positions of the output to obtain the predicted segmentation anomaly maps. The segmentation results for some categories in MVTEC dataset are shown in Fig. 1(More detailed anomaly segmentation results are shown in Appendix). We achieved a result that is close to the ground truth, demonstrating the ability of our method to segment unknown defects in high-resolution images.

Table 3. Detection performance of variants of CutX on the MVTEC dataset. From left to right are results of cutout, cutmix, cutpaste and our proposed method. The results are reported with AUROCs.

Category	Transforms			
	Cutout	**Cutpaste**	**Cutmix**	**Ours**
object	92.3	95.3	96.8	**97.4**
texture	97.3	98.4	98.7	**99.6**
Overall AVG	93.9	96.3	97.4	**98.1**

5 Conclusion

We design a finer-grained self-supervised proxy task, where the key lies in segmenting the augmented patches so as to simulate segmentation task in vision to extract the differences between normal and abnormal patterns. Also our approach does not require the simulated anomalies to match those in the real scenario, ensuring that the model obtains more robust decision boundary. Extensive experimental results confirm that SSAPS achieves excellent results in detecting surface defects in high-resolution images of industrial scenes. Furthermore, SSAPS also demonstrates good defect localization capabilities due to the pixel-level differential perception of normal and abnormal patterns learned by the model during the self-supervised training phase.

Acknowledgements The paper is supported by the Open Fund of Science and Technology on Parallel and Distributed Processing Laboratory under Grant WDZC20215250116.

References

1. Waleed Hilal, S Andrew Gadsden, and John Yawney. Financial fraud:: A review of anomaly detection techniques and recent advances. 2022.
2. Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019.
3. Chin-Chia Tsai, Tsung-Hsuan Wu, and Shang-Hong Lai. Multi-scale patch-based representation learning for image anomaly detection and segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3992–4000, 2022.
4. Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020.
5. Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pages 4393–4402. PMLR, 2018.
6. David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004.
7. Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Self-supervised out-of-distribution detection and localization with natural synthetic anomalies (nsa). *arXiv preprint arXiv:2109.15222*, 2021.
8. Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021.
9. Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020.

10. Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021.
11. Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, pages 01–06. IEEE, 2021.
12. Ye Zheng, Xiang Wang, Rui Deng, Tianpeng Bao, Rui Zhao, and Liwei Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. *arXiv preprint arXiv:2110.04538*, 2021.
13. Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022.
14. Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021.
15. N Cohen and Y Hoshen. Sub-image anomaly detection with deep pyramid correspondences. arxiv 2020. *arXiv preprint arXiv:2005.02357*.
16. Oliver Rippel, Patrick Mertens, and Dorit Merhof. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6726–6733. IEEE, 2021.
17. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
18. Shuang Mei, Hua Yang, and Zhouping Yin. An unsupervised-learning-based approach for automated defect inspection on textured surfaces. *IEEE Transactions on Instrumentation and Measurement*, 67(6):1266–1277, 2018.
19. Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018.
20. Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018.
21. Mohammad Sabokrou, Mohammad Khalooei, Mahmood Fathy, and Ehsan Adeli. Adversarially learned one-class classifier for novelty detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3379–3388, 2018.
22. Yufei Liang, Jiangning Zhang, Shiwei Zhao, Runze Wu, Yong Liu, and Shuwen Pan. Omni-frequency channel-selection representations for unsupervised anomaly detection. *arXiv preprint arXiv:2203.00259*, 2022.
23. Jonathan Pirnay and Keng Chai. Inpainting transformer for anomaly detection. In *International Conference on Image Analysis and Processing*, pages 394–406. Springer, 2022.
24. Minghui Yang, Peng Wu, Jing Liu, and Hui Feng. Memseg: A semi-supervised method for image surface defect detection using differences and commonalities. *arXiv preprint arXiv:2205.00908*, 2022.

25. Nikos Komodakis and Spyros Gidaris. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations (ICLR)*, 2018.
26. Izhak Golan and Ran El-Yaniv. Deep anomaly detection using geometric transformations. *Advances in neural information processing systems*, 31, 2018.
27. Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE international conference on computer vision*, pages 1422–1430, 2015.
28. Jouwon Song, Kyeongbo Kong, Ye-In Park, Seong-Gyun Kim, and Suk-Ju Kang. Anoseg: Anomaly segmentation network using self-supervised learning. *arXiv preprint arXiv:2110.03396*, 2021.
29. Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022.
30. Roy De Maesschalck, Delphine Jouan-Rimbaud, and Désiré L Massart. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50(1):1–18, 2000.
31. Jin-Hwa Kim, Do-Hyeong Kim, Saehoon Yi, and Taehoon Lee. Semi-orthogonal embedding for efficient unsupervised anomaly segmentation. *arXiv preprint arXiv:2105.14737*, 2021.
32. Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minho Jin, and Tomas Pfister. Learning and evaluating representations for deep one-class classification. *arXiv preprint arXiv:2011.02578*, 2020.
33. Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.