ORIGINAL PAPER

Self-supervised anomaly detection based on foreground enhancement and autoencoder reconstruction

Lijie Zhao[1] · Yuan Chai[1] · Qichun Zhang[2] · Hamid Reza Karimi[3]

Abstract

Anomaly detection algorithms typically utilize generative models to reconstruct anomaly regions. Post-processing is used to pinpoint the anomalies. However, the paucity of real-world anomaly samples and the complex image backgrounds pose significant challenges for anomaly detection. The work innovatively proposed a self-supervised anomaly detection method. An efficient channel attention mechanism in the autoencoder was introduced to improve the reconstruction performance. Besides, a foreground enhancement strategy was designed to distinguish the foreground from the background by maximizing the inter-class variance. The strategy reduced the effect of background noises and simulated various anomalies that were rare in real samples. The MVTecAD and BTAD datasets were used to experiment with anomaly detection and location. Experimental results demonstrated that our method achieved higher AUC and AP scores at both the image level and pixel level compared to other advanced methods. In particular, the average AP score increased by 12.5% at the pixel level.

Keywords Anomaly detection · Self-supervised method · Efficient channel attention · Foreground enhancement

1 Introduction

Anomaly detection algorithms are widely used thanks to their stable performance and high detection efficiency. One of the classical approaches focuses on reconstruction. This approach utilizes a neural network to encode and decode the input normal samples, with the reconstruction input as the training target. The distribution pattern of the normal samples can be learned, and anomalies are identified by analyzing the differences between the original and reconstructed images. Commonly used reconstruction-based methods are categorized into autoencoders (AE) [1–4] and generative adversarial networks (GAN) [5–7] according to the different training modes employed. Even frameworks [8, 9] that combine AE and GAN can achieve impressive results. Moreover, embedding-based methods [10, 11] have shown good performance in anomaly detection tasks. The basic principle is to match the features of test samples with normal samples. The inference step of such models involves a complex feature-matching process, which increases the computational cost of the model even if its training phase takes little time.

In practical applications, the scarcity and variety of real-world anomaly samples pose significant challenges to supervised learning. A self-supervised surface anomaly detection method is proposed to address the issue in the work. The main contributions are as follows.

- Our method based on autoencoder reconstruction introduced a novel combination of foreground enhancement strategy and efficient channel attention mechanism, which considerably improved anomaly detection performance.
- An anomaly generation module was designed to generate anomaly samples using a foreground enhancement strat-

✉ Qichun Zhang
  q.zhang17@Bradford.ac.uk

  Lijie Zhao
  zhaolijie@syuct.edu.cn

  Yuan Chai
  2021246@stu.syuct.edu.cn

  Hamid Reza Karimi
  hamidreza.karimi@polimi.it

1   College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, Liaoning, China

2   Department of Computer Science, University of Bradford, Bradford BD7 1DP, UK

3   Department of Mechanical Engineering, Politecnico di Milano, 32 20133 Milan, Italy

egy, which mitigated the impact of irrelevant background information on model learning.

- An efficient channel attention mechanism was introduced in the autoencoder. It captured cross-channel interaction information and enhanced the network's ability to extract channel features for better reconfiguration.
- Experimental validation on the MVTecAD [12] and BTAD [13] datasets demonstrated that our method outperformed other advanced approaches in multiple metrics. It achieved the highest improvement of up to 12.5%, especially in pixel-level average AP score.

2 Related work

Most surface anomaly detection models [7, 14, 15] aim to explore the broad patterns of normal samples. Only normal samples are used for reconstruction to train the model. MS-FCAE [16] is designed with multi-scale feature information based on AE, which provides different levels of contextual information for image reconstruction. Therefore, the reconstructed image is more accurate and clearer. AnoGAN [17] first introduces GAN into anomaly detection. However, it requires several iterations of optimization in the inference stage to generate a suitable normal image as a reference. The algorithm lacks the necessary computational efficiency to be deployed in real-time detection tasks. F-AnoGAN [18] introduces an additional encoder to extract image features based on GAN, which guides the generator to create the most matching images. In this regard, our work introduced an efficient channel attention mechanism [19] during the reconstruction of anomalous regions. This mechanism effectively captured inter-channel interactions and enhanced the capacity of the network to extract features, resulting in better reconstruction quality.

Some studies [20–22] seek to produce artificially simulated anomalous samples during the training phase to reveal the hidden differences between normal and anomalous samples. Specifically, CutPaste [20] utilizes augmentation techniques such as copy and paste to simulate anomalous samples by randomly copying a tiny rectangular region from the input image and pasting it onto the resulting image. DRAEM [21] creates anomalous areas by superimposing extra texture images as noise over normal images. Haselmann [22] adds rectangular masks at random to normal samples to simulate true anomalies. Considering the impact of background interference, we designed an effective foreground enhancement strategy. The strategy involved foreground extraction on the images and introduced noises to simulate anomaly generation, which resulted in more realistic anomaly samples for training.
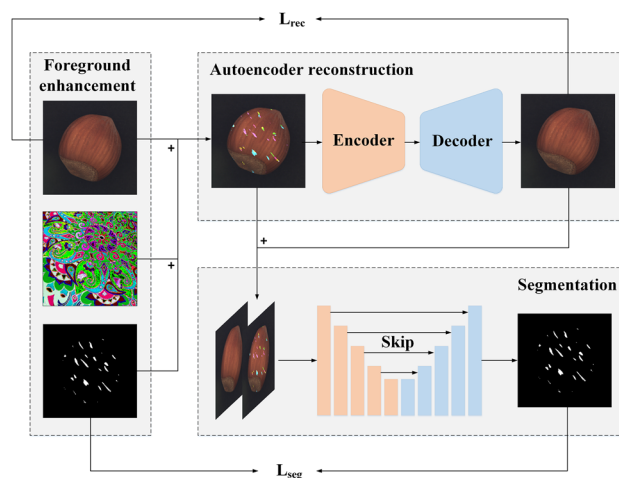

Fig. 1 The anomaly detection process of our method

3 Anomaly detection method

Our method consists of a foreground-enhanced anomaly generation module, an autoencoder reconstruction module and a segmentation module. The foreground-enhanced anomaly generation module is utilized to generate simulated anomaly samples by combining normal images with anomaly texture source images. The anomaly generation strategy can provide an arbitrary number of anomaly samples and accurate pixel-level segmentation ground truth maps, enabling our method to be trained without using real anomaly samples. The autoencoder with an efficient attention mechanism is trained using the L_{rec} to repair the anomalous regions. The input and output of autoencoder are concatenated and fed into the segmentation network, which is trained to localize anomalous regions using the L_{seg} (Fig. 1). A mean filter convolution layer is utilized to smooth the segmentation module's output. The anomaly score is calculated by selecting the maximum value from the smoothed anomaly score map.

3.1 Foreground enhancement

Anomalies appear in diverse manifestations within real-world scenarios, which poses challenges in comprehensive anomalies data collection. Consequently, the construction of ideal large-scale anomaly datasets for training supervised detection models becomes arduous. Therefore, an effective strategy is designed to simulate anomaly generation for self-supervised learning. Figure 2 depicts the anomaly generation strategy with foreground enhancement.

The noise image P is obtained by using a Perlin noise generator [23] to capture various anomalous shapes. It is binarized using a randomly uniformly sampled threshold T ($T = 0.5$) to form the anomaly mask image P_m. Besides, considering that in some actual collected image datasets,
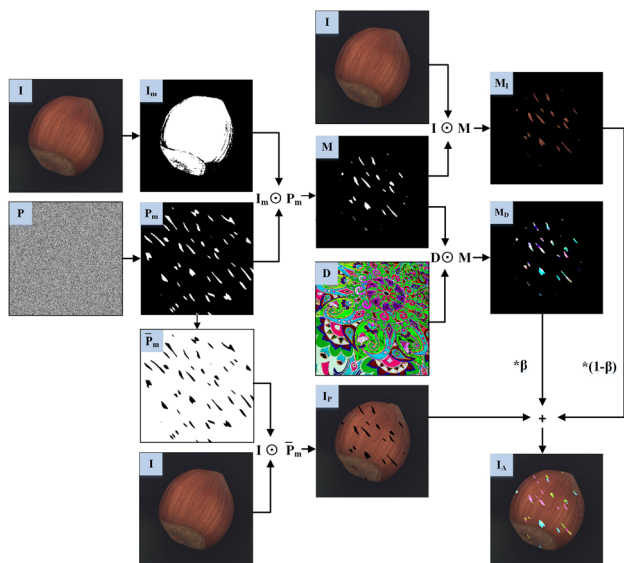
Fig. 2 Simulation anomaly generation strategy

Fig. 3 Multi-scale anomaly mask map

certain industrial components do not occupy a high enough proportion in the image, directly adding the anomaly mask could easily generate noise in the background. This increased disparity between the data distribution of real anomaly samples and simulated anomaly samples poses a greater challenge for the model to extract meaningful information. Therefore, a foreground enhancement strategy is applied to this kind of images. The Otsu method [24] is used to differentiate the foreground and background based on the maximization of inter-class variance. Original image I is then binarized to generate mask I_m. Element-wise multiplication is performed by multiplying masks P_m and I_m to obtain mask image M.

$$M = P_m \odot I_m \qquad (1)$$

where $\odot$ denotes the element-wise multiplication operation.

Anomaly texture source image D is drawn from a collection of anomalies that is not related to the distribution of the original image I. Anomaly image D is randomly enhanced with three methods chosen by the group: {sharpness, equalize, solarize, posterize, auto-contrast, brightness change}, which retains the diversity of anomaly. Enhanced texture image D and original image I are masked through mask M. Subsequently, they are blended with the original image I processed through mask $\bar{P}_m$ to obtain final simulated anomaly-generating image I_A.

$$I_A = (1 - \beta)(I \odot M) + \beta(D \odot M) + I \odot \bar{P}_m \qquad (2)$$

where β denotes the opacity parameter during blending, which is randomly and uniformly sampled from [0.1, 1.0], $\bar{P}_m$ is the inverse of P_m.
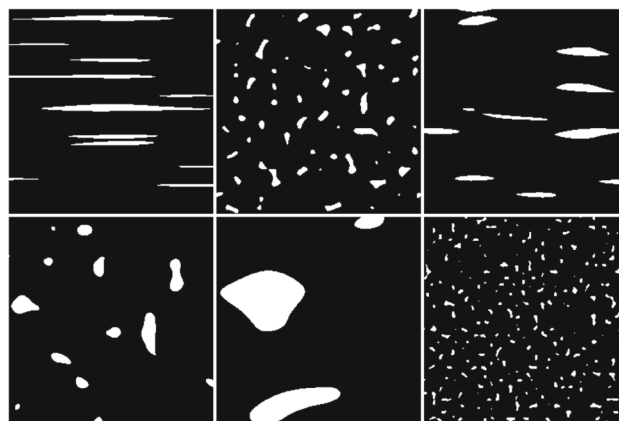
In addition to applying various data augmentations to anomaly texture source image D, there are other functions. This strategy randomly rotates 30% of input images I and Perlin noise P within $[-90°, 90°]$, which strengthens robustness. Furthermore, Perlin noise can randomly change the size during simulated anomaly generation by considering the diversity of various component anomalies in actual industrial environments. Thus, the granularity of the noise image is controlled to obtain anomaly mask image P_m with various sizes and shapes (Fig. 3).

3.2 Autoencoder reconstruction

An efficient channel attention (ECA) [19] mechanism, introduced in the encoding phase of the autoencoder, effectively captures cross-channel interaction information and enhances the network's feature extraction capability. The autoencoder can reconstruct the local anomalous pattern of input image I_A into a pattern closer to the normal sample distribution. Meanwhile, it maintains the non-anomalous areas of the original image unaltered and obtains reconstructed image I_r with the equal size to the original image. Figure 4 presents the architecture of the autoencoder.

ECA uses a dynamic convolution kernel to address the issue of extracting different range features for input feature maps with different numbers of channels. The convolution kernel adaptively changes its size through a function (Fig. 5).

$$k = \psi(C) = \left\| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right\|_{odd} \qquad (3)$$

where k is the convolution kernel size, C is the number of channels, $\| \quad \|_{odd}$ indicates that k can only be odd, and $\gamma = 2$ and $b = 1$ are used to change the ratio between the number of channels C and the convolution kernel size.
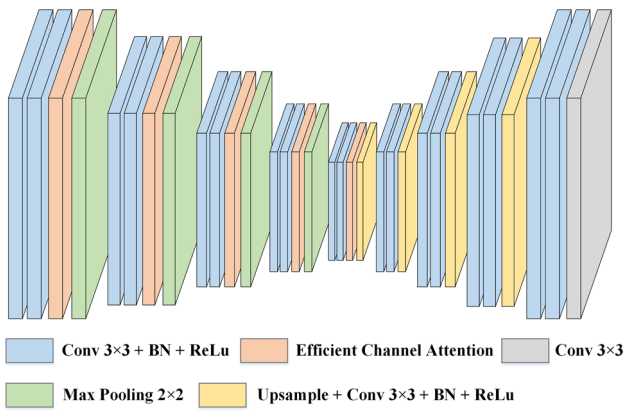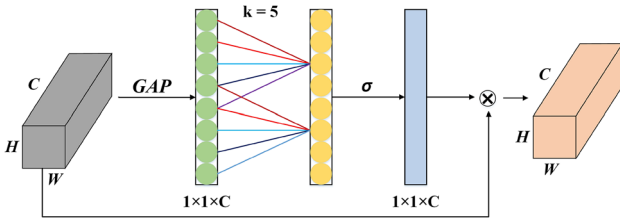
Fig. 4 Architecture of the autoencoder

Fig. 5 Process of efficient channel attention (ECA)

3.3 Segmentation

A U-net [25]-like structure is employed by the segmentation network. Input I_A and output I_r of the autoencoder are first concatenated along the channel dimension. They are then fed into the segmentation network to provide enough information for anomaly localization. Simultaneously, five downsampling convolution blocks are applied for multi-scale feature extraction. This part includes the original image with a total of 6 scales, which can fully extract features. The feature map of equivalent size from the feature extraction part is copied and fused at each stage of network upsampling. Eventually, the image is restored to its original size to obtain an accurate anomaly segmentation map. Figure 6 shows the architecture of the segmentation network.

3.4 Loss function

Structural similarity index mean (SSIM) [26] has become a common loss function in computer vision. It is typically utilized to evaluate the similarity of two images and considers three key features of an image (i.e., luminance, contrast and structure).

$$SSIM (x, y) = l (x, y)^{\alpha} \times c (x, y)^{\beta} \times s (x, y)^{\gamma} \qquad (4)$$

where $l (x, y)$, $c (x, y)$ and $s (x, y)$ are the luminance similarity, contrast similarity and structure similarity of images x
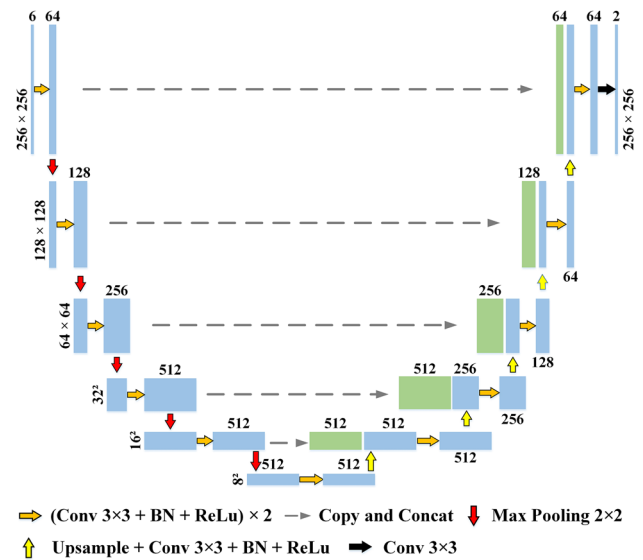

Fig. 6 Architecture of the segmentation network

and y, respectively. α, β and γ represent the balanced hyperparameters.

The L_2 loss is commonly utilized for anomaly detection algorithms, but adjacent pixels are assumed to be independent. Therefore, the SSIM loss is used enhance the interactivity between pixels additionally.

$$L_{SSIM} (I, I_r) = \frac{1}{N_P} \sum_{i=1}^{H} \sum_{j=1}^{W} 1 - SSIM (I, I_r)_{(i,j)} \qquad (5)$$

where H and W are the height and width of original image I, respectively. N_P is the number of pixels in I; I_r is the reconstructed image. $SSIM (I, I_r)_{(i,j)}$ is the SSIM value of I and I_r centered on coordinate (i, j), so the loss of reconstruction is defined by

$$L_{rec} (I, I_r) = \lambda L_{SSIM} (I, I_r) + L_2 (I, I_r) \qquad (6)$$

where λ is the balanced hyperparameter of two kinds of losses.

Focal loss [27] (L_{seg}) is applied to the segmentation network because it can solve imbalance between positive and negative samples and improve the robustness of accurate segmentation of complex samples. According to the network's reconstruction and segmentation goals, the overall loss during training is defined by

$$L (I, I_r, M_a, M) = L_{rec} (I, I_r) + L_{seg} (M_a, M) \qquad (7)$$

where I is the input image; I_r is the reconstructed image; M is the output segmentation mask; and M_a is ground truth.

4 Experiments

MVTecAD [12] and BTAD [13] datasets for anomaly detection and localization were tested to evaluate the effectiveness of our method. Our method was more targeted at anomaly detection in images with background interference. The MVTecAD and BTAD datasets contained different categories of anomaly images. However, the inclusion of image data without backgrounds (e.g., texture category images) was deemed irrelevant for demonstrating the efficacy of our methodologies. Therefore, not all categories in the two datasets were tested. The optimizer used in the training phase was Adam [28], with a total of 700 iterations. The initial learning rate was set to 0.0001 and decayed at the 560th and 630th iterations with a decay factor of 0.2. The input image size was uniformly scaled to 256×256, and the input batch size was 16.

A series of evaluation standards were calculated to quantitatively assess the detection capacity. The area under the receiver operating characteristic (AUROC) curve was the primary metric to compare the anomaly detection results. However, most of the anomalous areas were relatively small in practical applications. The metric value was affected by a large number of non-anomalous pixels while only a small number of anomalous pixels were detected. As a result, the pixel-level AUROC did not reflect the localization accuracy well. Therefore, the work additionally calculated the average precision (AP) metric, which represented the region under the curve of precision and recall rates. It was particularly well suited for highly imbalanced categories, notably in anomaly detection scenarios with precision playing a pivotal role.

4.1 Comparison with existing methods

Our approach is compared with unsupervised anomaly detection methods for images developed in recent years, including GANomaly [15], PaDiM [11], STAD [29], CutPaste [20], and DRAEM [21]. GANomaly combines autoencoders and generative adversarial networks; PaDiM extracts patch embeddings from the input image using a previously trained CNN; STAD solves the unsupervised anomaly segmentation problem using a student–teacher network; and CutPaste and DRAEM attempt to generate simulated anomalous samples during training. In summary, our method outperforms other methods and achieves the highest AUC and AP scores at both the image level and pixel level.

Table 1 displays the image-level AUROC metric. Our method achieves the highest or second-highest AUROC scores in each of the five categories in the MVTecAD dataset. Compared to the advanced method DRAEM, our method further improved the average image-level AUROC score by 0.3% across. This improvement is evident not only on the screw dataset, where anomalous regions are minimal and

Table 1 Comparative results of image-level AUROC (%) on the MVTecAD dataset

Class	GANomaly	CutPaste	DRAEM	PaDiM	Ours
Hazelnut	87.4	92.0	**100**	92.0	99.7
Pill	67.1	86.1	**94.3**	93.3	92.4
Screw	**100**	81.3	94.1	85.8	98.1
Toothbrush	70.0	100	**100**	96.1	**100**
Zipper	74.4	97.9	**100**	90.3	**100**
Average	79.8	93.5	97.7	91.5	**98.0**

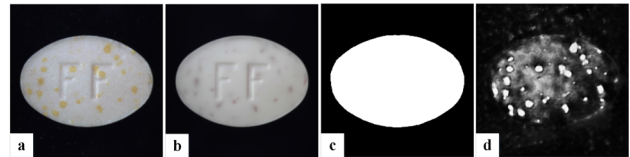Bold numbers indicate the maximum value of the performances

Fig. 7 Anomaly case analysis. **a** Anomaly image; **b** Reconstructed image; **c** Ground truth; and **d** Detect result

challenging to distinguish, but also on the toothbrush dataset with limited training samples, highlighting the effectiveness of our approach.

The pixel-level AUROC and AP metrics in Table 2 show the excellent performance of our method. The average AUROC is improved by 0.6% over PaDiM, and the average AP is significantly improved by 12.5% over DRAEM. These improvements can be owing to the ECA mechanism, which enhances the model's reconstruction ability for images containing irregular workpiece anomalies. Additionally, the foreground enhancement strategy eliminates background interferences, which allows the model to acquire more valuable information.

The pill dataset shows the poorest detection performance. The original training samples of the pills all have spots of the same color. However, several samples only have different spot colors and no surface anomalies during testing. The model mistakenly identifies this type of anomaly as a staining anomaly rather than a category anomaly in this case. It leads to significant disparities between the segmentation result and the ground truth and affects the detection metric (Fig. 7).

To fully demonstrate our advantages, the same settings are used as the MVTecAD dataset to test the more challenging BDAD dataset without any data augmentation. The anomaly detection results are compared with traditional algorithms [13, 21, 30]. As shown in Table 3, our method outperforms other advanced algorithms in terms of average AUROC scores at the image level and pixel level for two categories within the BTAD dataset, demonstrating exceptional efficacy.

Figure 8 shows the visualization of anomaly detection results, and each column displays the anomalous image,

Table 2 Comparative results of pixel-level AUROC (%) and AP (%) on the MVTecAD dataset

Class	STAD		DRAEM		PaDiM		Ours	
	AUROC	AP	AUROC	AP	AUROC	AP	AUROC	AP
Hazelnut	98.2	57.8	99.3	82.4	98.1	61.1	**99.6**	**87.2**
Pill	96.5	62.0	**97.2**	42.3	95.7	61.2	95.3	**63.7**
Screw	97.4	7.8	99.1	62.0	98.4	21.7	**99.7**	**68.2**
Toothbrush	97.9	37.7	97.4	46.1	98.8	54.7	**99.0**	**69.5**
Zipper	95.6	36.1	98.5	72.1	98.4	58.2	**98.9**	**78.8**
Average	97.1	40.3	98.3	61.0	97.9	51.4	**98.5**	**73.5**

Bold numbers indicate the maximum value of the performances

Table 3 Comparative results of image-level and pixel-level AUROC (%) on the BTAD dataset

Class	P-SVDD		DRAEM		VT-ADL		Ours	
	Image level	Pixel level	Image level	Pixel level	Image level	Pixel level	Image level	Pixel level
01	95.7	91.6	95.2	88.6	**97.6**	**99.0**	97.3	90.9
02	82.1	91.0	99.4	87.6	82.6	77.0	**99.8**	**91.8**
Average	88.9	91.3	97.3	88.1	90.1	88.0	**98.6**	**91.4**

Bold numbers indicate the maximum value of the performances

Fig. 8 Visualization of anomaly detection results

Table 4 Ablation results on MVTecAD dataset

Foreground Enhancement	ECA Module	Image level AUROC %	Pixel level AUROC %	Pixel level AP %
		96.7	97.0	59.9
✓		97.3	97.7	67.7
	✓	97.0	97.5	62.5
✓	✓	**98.0**	**98.5**	**73.5**

Bold numbers indicate the maximum value of the performances

Table 5 Ablation results on BTAD dataset

Foreground Enhancement	ECA Module	Image level AUROC %	Pixel level AUROC %
		91.7	79.4
✓		98.2	88.0
	✓	97.3	87.7
✓	✓	**98.6**	**91.4**
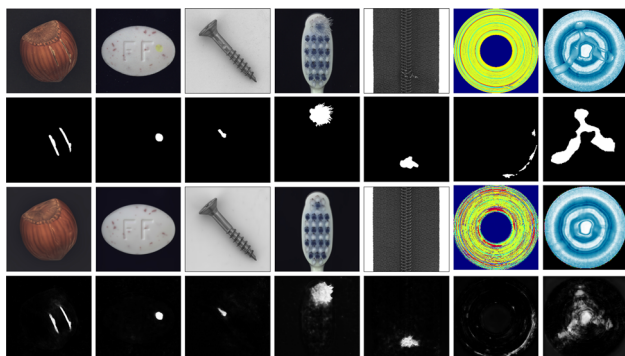
Bold numbers indicate the maximum value of the performances

ground truth, reconstructed image, and detection result in sequential order. It can be observed that our algorithm can clearly reconstruct anomalous images while accurately locating surface anomalies on the products.

4.2 Ablation study

To validate the effectiveness of our method, ablation studies were carried out by applying the foreground enhancement strategy and adding multiple ECA modules to the baseline model. The baseline model directly simulated anomalies by adding noises to the input image without any enhancement strategy. Additionally, the network did not contain any attention module. Tables 4 and 5 show the ablation results for each dataset. By comparing, it is evident that both modules show improvements compared to the baseline model. Our method significantly improves the average pixel-level AP scores by 13.6% in the MVTecAD dataset when both modules are used

simultaneously. Similarly, our method notably increases the average image-level and pixel-level AUROC scores by 6.9% and 12% in the BTAD dataset, respectively. The ablation results further demonstrate the effectiveness of each module.

In Fig. 9, each group presents the input anomaly image as well as the localization result of the baseline model, the model with only added ECA modules, the model only using foreground enhancement strategy, and the model combining foreground enhancement strategy and ECA modules. It is evident that the combination of the foreground enhancement strategy and the ECA modules yields the best anomaly localization results.
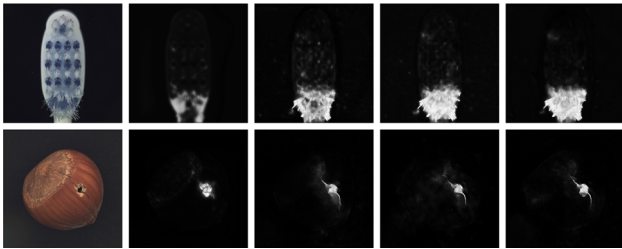
Fig. 9 Comparison of ablation results

5 Conclusion

A self-supervised method for surface anomaly detection was proposed in this paper. The method only required normal samples during training and simulated real anomalies through an anomaly generation strategy with the foreground enhancement. It could mitigate the impact of invalid information in the background on model learning to some extent. Accurate anomaly detection results could be obtained through an autoencoder with efficient channel attention and a U-net-like network for fine segmentation, addressing the issue of imprecise anomaly localization in existing methods. Experimental results on the MVTecAD and BTAD datasets demonstrated that our method achieved excellent performance in anomaly detection and localization. In particular, compared with other advanced methods, the pixel-level average AP was significantly improved by 12.5%. The proposed method provided a better depiction of anomaly segmentation details and exhibits superior overall detection performance.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. All other used public datasets are available and cited in reference.

Declarations

Conflict of interest The authors have no relevant financial or nonfinancial interests to disclose.

Ethics approval and consent to participate Not applicable.

Consent for publication All authors agreed on the final approval of the version to be published.

References

1. 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)
2. Chandrakala, S., Shalmiya, P., Srinivas, V., Deepak, K.: Object-centric and memory-guided network-based normality modeling for video anomaly detection. Signal Image Video Process. **16**(7), 2001–2007 (2022)
3. Jiang, R., Xue, Y., Zou, D.: Interpretability-aware industrial anomaly detection using autoencoders. IEEE Access (2023)
4. Hu, X., Lian, J., Zhang, D., Gao, X., Jiang, L., Chen, W.: Video anomaly detection based on 3d convolutional auto-encoder. Signal Image Video Process. **16**(7), 1885–1893 (2022)
5. Xu, C., Ni, D., Wang, B., Wu, M., Gan, H.: Two-stage anomaly detection for positive samples and small samples based on generative adversarial networks. Multimed. Tools Appl. **82**(13), 20197–20214 (2023)
6. Wang, W., Chang, F., Liu, C.: Mutuality-oriented reconstruction and prediction hybrid network for video anomaly detection. SIViP **16**(7), 1747–1754 (2022)
7. Zenati, H., Foo, C.S., Lecouat, B., Manek, G., Chandrasekhar, V.R.: Efficient Gan-based anomaly detection. arXiv preprint arXiv:1802.06222 (2018)
8. Li, X., Jing, J., Bao, J., Lu, P., Xie, Y., An, Y.: Otb-aae: Semi-supervised anomaly detection on industrial images based on adversarial autoencoder with output-turn-back structure. IEEE Trans. Instrum. Meas. (2023)
9. Luo, Y., Ma, Y.: Anomaly detection for image data based on data distribution and reconstruction. Appl. Intell. 1–11 (2023)
10. 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)
11. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV, pp. 475–489 (2021)
12. 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)
13. Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., Foresti, G.L.: Vt-adl: a vision transformer network for image anomaly detection and localization. In: 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), pp. 01–06 (2021)
14. Tang, T.-W., Kuo, W.-H., Lan, J.-H., Ding, C.-F., Hsu, H., Young, H.-T.: Anomaly detection neural network with dual auto-encoders

Gan and its industrial inspection applications. Sensors **20**(12), 3336 (2020)

15. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: semi-supervised anomaly detection via adversarial training. In: Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14, pp. 622–637 (2019)

16. Yang, H., Chen, Y., Song, K., Yin, Z.: Multiscale feature-clustering-based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects. IEEE Trans. Autom. Sci. Eng. **16**(3), 1450–1467 (2019)

17. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25–30, 2017, Proceedings, pp. 146–157 (2017)

18. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-anogan: fast unsupervised anomaly detection with generative adversarial networks. Med. Image Anal. **54**, 30–44 (2019)

19. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542 (2020)

20. Li, C.-L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9664–9674 (2021)

21. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem—a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8330–8339 (2021)

22. Haselmann, M., Gruber, D.P., Tabatabai, P.: Anomaly detection using deep learning based image completion. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1237–1242 (2018)

23. Perlin, K.: An image synthesizer. ACM Siggraph Comput. Graph. **19**(3), 287–296 (1985)

24. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)

25. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pp. 234–241 (2015)

26. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

27. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)

28. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

29. 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)

30. Yi, J., Yoon, S.: Patch svdd: patch-level svdd for anomaly detection and segmentation. In: Proceedings of the Asian Conference on Computer Vision (2020)