IPAD:Inpainting in Anomaly Detection

Jiaxiang Wang[1*], Haodi Xu[2*], Yinhao Liu[3*], Hangyang Kong[4*], Yinuo Zhang[5*]

[1]36920231153234 class for Institute of AI
[2]23320231154455 class for Institute of Information
[3]36920231153218 class for Institute of AI
[4]36920231153202 class for Institute of AI
[5]36920231153264 class for Institute of AI

Abstract

Anomaly detection is an important application in large-scale industrial manufacturing. And Reconstruction-based methods play an important role in unsupervised anomaly detection in images.We introduce an In-painting method for Anomaly Detection and propose a novel approach to constructing pseudo-anomalous images.Our method learns a joint representation of anomalous images and their anomaly-free reconstructions, while simultaneously learning a decision boundary between normal and anomalous examples.On the challenging Visa anomaly detection dataset, we get good performance.

Introduction

Image anomaly detection and localization task aims to identify abnormal images and locate abnormal subregions. The technique to detect the various anomalies of interest has a broad set of applications in industrial inspection(Bergmann et al. 2019a)(Defard et al. 2021).

In industrial scenarios, anomaly detection and localization is especially hard, as abnormal samples are scarce and anomalies can vary from subtle changes such as thin scratches to large structural defects, e.g. missing parts. Some examples from the MVTec AD benchmark (Bergmann et al. 2019a) along with results from our proposed method are shown in Figure 1.This situation prohibits the supervised methods from approaching.

Reconstructive methods, such as Autoencoders (Bergmann et al. 2019b)(Akcay, Atapour-Abarghouei, and Breckon 2019)(Tang et al. 2020) and GANs(Schlegl et al. 2017)(Schlegl et al. 2019)], have been extensively explored since they enable learning of a powerful reconstruction subspace, using only anomaly-free images. Relying on poor reconstruction capability of anomalous regions, not observed in training, the anomalies can then be detected by thresholding the difference between the input image and its reconstruction.

However, the traditional reconstruction method is still challenging when it is determined whether there is no significant difference between the existence and the normal appearance.Recent research has taken into account the differences between the network extracted from the general network and the network specially used for the network without abnormal images to improve the discernic.At the same time, some dense clusters focusing on non -abnormal texture in the deep space to prevent abnormalities from being mapped to the position close to the abnormal sample.

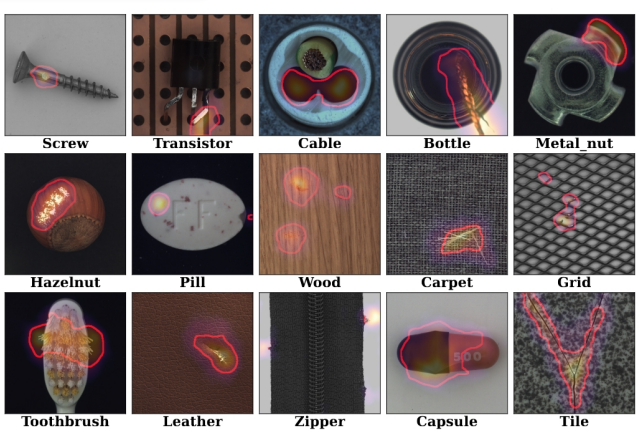

Figure 1: Visualization of samples in MVTec AD. The produced anomaly maps superimposed on the images. Anomaly region of high anomaly score is colored with orange. The red boundary denotes contours of actual segmentation maps for anomalies.

One of the disadvantages of generating methods is that they only learn models from non -abnormal data, and they cannot explicitly optimize the discriminator abnormal detection, because they cannot obtain positive examples (that is, abnormal) during training.Although the use of synthetic abnormalities can be considered to train the discriminating division method, this often leads to excessive fitting of the synthetic appearance, and it makes it difficult for the learning boundary to be learned to be promoted to real abnormalities.

In order to solve this problem, we consider a type of generating model -diffusion model (DM), which has achieved significant success in image generation.The diffusion model adds noise to the image by iteratively, and then iterates the noise, thereby realizing the mapping of the image to a specific flow.We believe that we can use the diffusion model to

*These authors contributed equally.

learn the characteristics of flow mapping and use it for unsupervised external detection.By reinforcing the image from the original current and using the training model for repair and mapping, we can measure the distance between the image and the original image, and the image outside the detection domain is based on this distance.

We propose, as our main contribution, a new deep surface anomaly detection network, discriminatively trained in an end-to-end manner on synthetically generated just-out-of-distribution patterns, which do not have to faithfully represent the target-domain anomalies.

In this paper,the network is composed of a reconstructive subnetwork, followed by a discriminative sub-network.The reconstructive sub-network is a diffusion model trained on the in-domain data is repaired to map the elevated image, while the discriminative subnetwork learns a discriminative model over the joint appearance of the original and reconstructed images, producing a high-fidelity per-pixel anomaly detection map.

Overall, our study proposes two unique approaches to improve the performance of image anomaly detection and unsupervised extraterritorial detection. By improving the reconstructed subnetwork of the deep surface anomaly detection network, we avoid the overfitting problem of synthetic appearance and improve the generalization ability of the model At the same time, by introducing the diffusion model for extraterritorial detection, we take advantage of the diffusion model to learn manifold mapping and realize unsupervised extraterritorial detection. These two methods complement each other and bring new possibilities to the field of anomaly detection.

In this work, we explore a novel approach, in which brings new possibilities to the field of anomaly detection. Our contributions are as follows

- We introduce an effective in-painting model to reconstruct more real images,which is significant for segmentation.
- We propose a novel method to generate pseudo-anomalous images,and it is close to real anomalous images
- We get good performance in both image-auroc and pixel-auroc in the VisA dataset

Related Work

Many surface anomaly detection methods focus on image reconstruction and detect anomalies based on image reconstruction error.Autoencoders (Zhou and Paffenroth 2017)(Kingma and Welling 2019)are trained on data of healthy subjects. Any deviations from the learned distribution then lead to a high anomaly score.Other approaches focus on Generative Adversarial Networks (GANs) (Goodfellow et al. 2017) for image-to-image translation(Baumgartner et al. 2017).

However, training of GANs is challenging and requires a lot of hyperparameter tuning. Furthermore, additional loss terms and changes to the architecture are required to ensure cycle-consistent results.In (Schlegl et al. 2019)(Schlegl et al. 2017), a GAN (Goodfellow et al. 2019) is trained to generate images that fit the training distribution. In (Schlegl et al. 2019) an encoder network is additionally trained that finds the latent representation of the input image that minimizes the reconstruction loss when used as the input by the pre-trained generator. The anomaly score is then based on the reconstruction quality and the discriminator output. In (Wu et al. 2020)an interpolation auto-encoder is trained to learn a dense representation space of in-distribution samples. The anomaly score is then based on a discriminator, trained to estimate the distance between the input-input and input-output joint distributions, however the approach to surface anomaly detection remains generative as the discriminator evaluates the reconstruction quality.

Instead of the commonly used image space reconstruction, the reconstruction of pretrained network features can also be used for surface anomaly detection (Bergmann et al. 2019a)(Shi, Yang, and Qi 2021). Anomalies are detected based on the assumption that features of a pre-trained network will not be faithfully reconstructued by another network trained only on anomaly-free images. Alternatively (Defard et al. 2021) propose surface anomaly detection as identifying significant deviations from a Gaussian fitted to anomaly-free features of a pre-trained network.This requires a unimodal distribution of the anomaly-free visual features which is problematic on diverse datasets.

Recently, a class of generative models – the diffusion models (DM)(Sohl-Dickstein et al. 2015)(Ho et al. 2022)– have gained increasing popularity. DMs formulate two processes: The forward process converts an image to a sample drawn from a noise distribution by iteratively adding noise to its pixels; the backward process maps a noise image towards a specific image manifold by iteratively removing noise from the image. A dedicated neural network is trained to perform the denoising steps in the backward process.

Proposed method

The proposed discriminative joint reconstructionanomaly embedding method is composed from an in-painting and a discriminative sub-networks. The reconstructive subnetwork is trained to implicitly detect and reconstruct anomalies, employing semantically plausible anomaly-free content while preserving the non-anomalous regions of the input image. Concurrently, the discriminative sub-network learns a joint reconstruction-anomaly embedding and generates accurate anomaly segmentation maps from the concatenated reconstructed and original appearances.. And we propose a new way to to constructing pseudo-anomalous imagewhich can generate images that are closer to real anomalies.

An New Anomaly Simulation Strategy

In traditional reconstruction methods, artificial anomalies are added to the entire image to generate pseudo-anomalous images, which are then subjected to image reconstruction. However, we know that anomalies only occur on the surface of objects and do not manifest in the background image. Adding noise to the background is meaningless. Therefore,

we extract the foreground of the image, add noise only to the surface of objects, thereby generating pseudo-anomalous images that better align with real scenarios. This process leads to improved reconstruction training. We show the

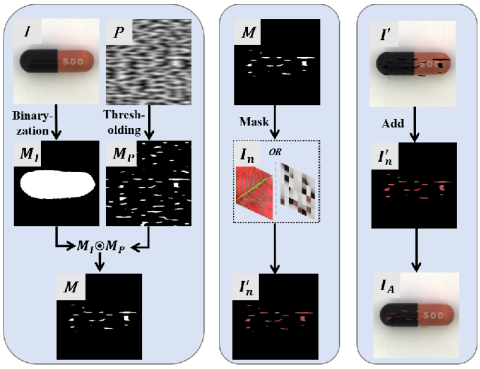

Figure 2: The three steps to add noise to generate pseudo-anomalous images

method steps in Fig.2. First, generate two-dimensional Perlin noise, and then obtain a mask through binarization with a threshold T. Considering that some industrial components in the image are relatively small in proportion, directly processing data augmentation can easily introduce noise in the background of the image. Therefore, we adopt a foreground enhancement strategy for such images. That is, binarize the input image to obtain a mask and use opening or similar operations to remove noise generated during the binarization process. Subsequently, the final mask image is obtained by element-wise multiplication of the two obtained masks. Then, perform element-wise multiplication of the mask image and the noise image, introducing a transparency factor in this process to balance the fusion of the original image and the noise image, making the simulated anomaly patterns closer to real anomalies. Finally, invert the mask image M, and then perform element-wise multiplication with the original image I to obtain the image I', resulting in the final simulated anomalous image.

In-painting Model Reconstruction

The reconstruction sub-network uses in-painting method.we extend the image-level filtering to the deep feature level and propose the semantic filtering, which can complete large missing areas but loses details.To address the issues, we propose a novel filtering technique, Multi-level Interactive Siamese Filtering (MISF), which contains two branches: kernel prediction branch (KPB) and semantic image filtering branch (SIFB). These two branches are interactively linked at semantic pixel levels. SIFB provides multi-level features for KPB while KPB predicts dynamic kernels for SIFB. MISF can utilize the smoothness prior across neighbors explicitly and reconstruct clean pixels or features by linearly combining the neighbors.

$$I' = I \odot K \tag{1}$$

where$I \in R^{W\times H}$ is the corrupted image and $I' \in R^{W\times H}$is the completed counterpart. The tensor $K \in R^{W\times H\times N^2}$ contains HW kernels for filtering all pixels. The operation $\odot$ denotes the pixel-wise filtering. We can expand the above equation as

$$I'[p] = \sum_{q\epsilon N_p} K_p[q-p]I[q] \tag{2}$$

Here, p and q are the coordinates of pixels in the image while the set N_p contains N^2 neighboring pixels of p. The matrix $K_p \in R^{N\times N}$is the pth vector of K and determines the weights for all pixels in N_p, which is also known as the kernel for the pixel p Predictive filtering is a widely used image restoration technique and can address image denoising tasks. Here, we formulate the image inpainting as the pixel-wise predictive filtering task.For image inpainting, the pixels at the boundary of missing areas are reasoned by their neighboring pixels. The principle is that the missing pixels do not break the local structure. Meanwhile, the related pixels can be used to reconstruct the missing pixels. However, the local structures around missing pixels are diverse and may distinguish them from each other. To adapt the context variations, we can train a predictive network to estimate the kernels for all pixels according to the input image.

$$K = \varphi(I) \tag{3}$$

Our semantic filtering is an improved encoder-decoder network that contains an extra 'dynamic convolution layer.We show the framework in Fig.3 . MISF further makes the dynamic process conditional on the multi-level features. As a result, the parameters of the dynamic convolution are element-wise and dynamically tuned according to different images and their semantic meaning through the predictive network. The advantages of dynamic convolution have been evidenced in many works. However, these works mainly focus on the image classification task. They predict convolutional parameters dynamically, according to the input features. In contrast, our work presents the importance of dynamic convolution for image inpainting and predicts dynamic convolutional parameters based on the raw input and deep features jointly with an element-wise way.

Discriminative sub-network

The discriminative sub-network uses U-Net [21]-like architecture. The sub-network input Ic is defined as the channel-wise concatenation of the reconstructive subnetwork output Ir and the input image I. Due to the normality-restoring property of the reconstructive subnetwork, the joint appearance of I and Ir differs significantly in anomalous images, providing the information necessary for anomaly segmentation.The discriminative sub-network learns the appropriate distance measure automatically. The network outputs an anomaly score map Mo of the same size as I. Focal Loss[14] (Lseg) is applied on the discriminative subnetwork output to increase robustness towards accurate segmentation of hard examples.

Experiments

Experimental Setup

Dataset The VisA dataset contains 12 subsets corresponding to 12 different objects. Figure 5 gives images in VisA.

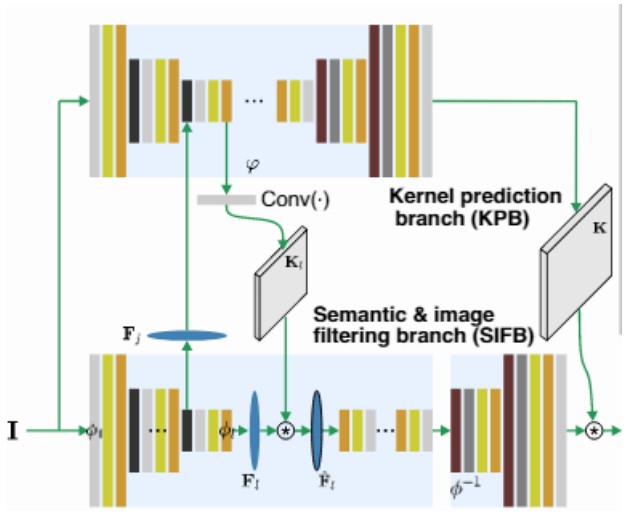

Figure 3: The overview of inpainting model.SIFB provides multi-level features for KPB while KPB predicts dynamic kernels for SIFB

There are 10,821 images with 9,621 normal and 1,200 anomalous samples. Four subsets are different types of printed circuit boards (PCB) with relatively complex structures containing transistors, capacitors, chips, etc. For the case of multiple instances in a view, we collect four subsets: Capsules, Candles, Macaroni1 and Macaroni2. Instances in Capsules and Macaroni2 largely differ in locations and poses. Moreover, we collect four subsets including Cashew, Chewing gum, Fryum and Pipe fryum, where objects are roughly aligned. The anomalous images contain various flaws, including surface defects such as scratches, dents, color spots or crack, and structural defects like misplacement or missing parts. There are 5–20 images per defect type and an image may contain multiple defects. The defects were manually generated to produce realistic anomalies. All images were acquired using a high-resolution RGB sensor. Both image and pixel-level annotations are provided.

Evaluation Metrics We used the area under the receiver operator curve (AUROC) based on produced anomaly scores to calculate anomaly detection at image-level performance (AUROC sample). Localization performance was evaluated using the AUROC at pixel-level.

Experiments results

We present our results in three categories:anomaly detection,anomaly localization,qualitative results.

Anomaly Detection We extensively compare our method with those published methods in the past two years.The comparison results on Visa are shown in Table 1.For a fair comparison, we reproduce all these methods with the same backbone as in our model. Thus, despite using the unmodified code from the official repositories, we are not able to exactly reproduce the original results, but our numbers are very close.We get good performance in image-auroc and 0.5% higher than NSA(ECCV 2022) and 7.6% higher than PyramidFlow(CVPR 2023). Compare to SOTA, we slightly lower about 3% to the RD++(CVPR 2023).

	NSA (ECCV2022)	PyramidFlow (CVPR2023)	RD++ (CVPR2023)	Ours
candle	90.1	78.9	95.6	95.9
capsules	87.6	81.9	89.0	95.0
cashew	92.7	93.7	98.2	94.7
chewinggum	96.0	87.5	98.3	90.2
fryum	90.3	83.9	95.3	88.6
macaroni1	95.6	80.3	93.6	95.2
macaroni2	71.8	76.4	83.0	85.3
pcb1	93.5	89.9	96.8	89.6
pcb2	97.7	88.7	95.8	92.1
pcb3	93.5	78.6	96.8	93.0
pcb4	96.6	89.6	99.6	96.7
pipe_fryum	94.2	84.0	99.7	88.5
average	91.6	84.5	95.1	92.1

Table 1: Detailed image-level AUROC on the Visa dataset

	NSA (ECCV2022)	PyramidFlow (CVPR2023)	RD++ (CVPR2023)	Ours
candle	97.8	75.4	98.3	92.6
capsules	84.4	95.8	99.3	96.9
cashew	85.5	94.6	94.0	91.7
chewinggum	98.5	95.3	98.0	97.6
fryum	80.5	93.7	96.7	89.4
macaroni1	85.9	95.4	99.7	95.0
macaroni2	76.0	94.1	98.2	98.6
pcb1	84.5	97.3	99.8	96.8
pcb2	94.1	96.9	98.8	90.7
pcb3	93.3	97.5	99.3	85.3
pcb4	96.7	90.1	98.5	91.1
pipe_fryum	97.5	97.4	99.0	97.8
average	89.6	93.6	98.3	93.6

Table 2: Detailed pixel-level AUROC on the Visa dataset

Anomaly Localization Our method can achieve significantly better results than some methods depending on patch-wise discrepancy.Note that the results in Table 2 show that our method can achieve much better results than NSA and equal to Pyramid when using the same backbone. What's more, in addition to the macaroni2 class, our method achieves best performance in these methods.

Qualitative Results The following is the visualization of our results Fig4. It can be seen that our method can better locate the area where the anomaly is located, which is very close to groundtruth, which provides convenience for the anomaly detection of our industrial components.

Conclusion

We propose a new algorithm for industrial anomaly detection and localization with reconstruction sub-network and the discriminative sub-network.We introduce inpainting method to reconstruct more real.In addition, We use a novel method to generate pseudo-anomalous images. Our model gets good performance and this is a remarkable result since our model is not trained on real anomalies. We find that an accurate decision boundary can be well estimated by learning the extent of deviation from reconstruction on simple simulations rather than learning either the normality or real anomaly appearance.However,It does not perform well in

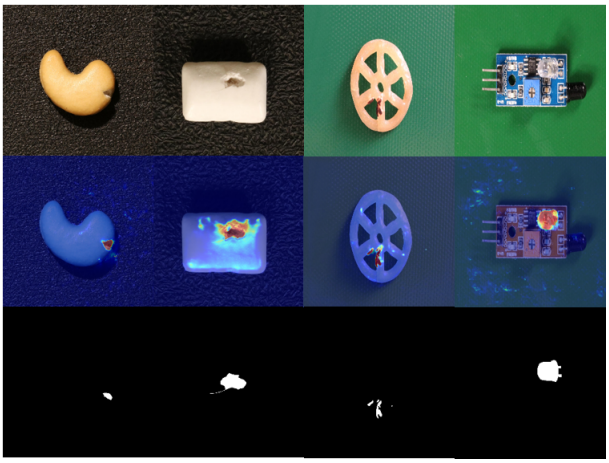

Figure 4: Anomalies in Visa from left to right: ”cashew”, ”chewinggum”, ”fryum”,”pcb3”

the case of missing parts,we will focus on this problem and look for more efficient segmentation methods

References

Akcay, S.; Atapour-Abarghouei, A.; and Breckon, T. P. 2019. *GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training*, 622–637.

Baumgartner, C.; Koch, L.; Tezcan, K.; Ang, J.; and Konukoglu, E. 2017. Visual Feature Attribution using Wasserstein GANs. *Cornell University - arXiv,Cornell University - arXiv*.

Bergmann, P.; Fauser, M.; Sattlegger, D.; and Steger, C. 2019a. MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Bergmann, P.; Löwe, S.; Fauser, M.; Sattlegger, D.; and Steger, C. 2019b. Improving Unsupervised Defect Segmentation by Applying Structural Similarity to Autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*.

Defard, T.; Setkov, A.; Loesch, A.; and Audigier, R. 2021. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. *Cornell University - arXiv,Cornell University - arXiv*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, D., Yoshua. 2019. Generative Adversarial Nets.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2017. GANGenerative Adversarial Nets. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 177–177.

Ho, J.; Jain, A.; Abbeel, P.; and Berkeley. 2022. Denoising Diffusion Probabilistic Models.

Kingma, D. P.; and Welling, M. 2019. An Introduction to Variational Autoencoders.

Schlegl, T.; Seeböck, P.; Waldstein, S. M.; Langs, G.; and Schmidt-Erfurth, U. 2019. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 30–44.

Schlegl, T.; Seeböck, P.; Waldstein, S. M.; Schmidt-Erfurth, U.; and Langs, G. 2017. *Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery*, 146–157.

Shi, Y.; Yang, J.; and Qi, Z. 2021. Unsupervised anomaly segmentation via deep feature reconstruction. *Neurocomputing*, 9–22.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv: Learning,arXiv: Learning*.

Tang, T.-W.; Kuo, W.-H.; Lan, J.-H.; Ding, C.-F.; Hsu, H.; and Young, H.-T. 2020. Anomaly Detection Neural Network with Dual Auto-Encoders GAN and Its Industrial Inspection Applications. *Sensors*, 3336.

Wu, Y.; Balaji, Y.; Vinzamuri, B.; and Feizi, S. 2020. Mirrored Autoencoders with Simplex Interpolation for Unsupervised Anomaly Detection. *arXiv e-prints,arXiv e-prints*.

Zhou, C.; and Paffenroth, R. C. 2017. Anomaly Detection with Robust Deep Autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.