



---

*Research article*

## Weakly supervised anomaly detection based on sparsity prior

Kaixuan Wang<sup>†</sup>, Shixiong Zhang<sup>†</sup>, Yang Cao and Lu Yang\*

School of Automation Engineering, University of Electronic Science and Technology of China, Sichuan, China

<sup>†</sup> The authors contributed equally to this work.

\* **Correspondence:** Email: yanglu@uestc.edu.cn.

**Abstract:** Anomaly detection is a binary classification task, which is to determine whether each pixel is abnormal or not. The difficulties are that it is hard to obtain abnormal samples and predict the shape of abnormal regions. Due to these difficulties, traditional supervised segmentation methods fail. The usual weakly supervised segmentation methods will use artificially generate defects to construct training samples. However, the model will be overfitted to artificially generated defects during training, resulting in insufficient generalization ability of the model. In this paper, we presented a novel reconstruction-based weakly supervised method for sparse anomaly detection. We proposed to use generative adversarial networks (GAN) to learn the distribution of positive samples, and reconstructed negative samples which contained the sparse defect into positive ones. Due to the nature of GAN, the training dataset only needs to contain normal samples. Subsequently, the segmentation network performs progressive feature fusion on reconstructed and original samples to complete the anomaly detection. Specifically, we designed the loss function based on kullback-leibler divergence for sparse anomalous defects. The final weakly-supervised segmentation network only assumes a sparsity prior of the defect region; thus, it can circumvent the detailed semantic labels and alleviate the potential overfitting problem. We compared our method with the state of the art generation-based generative anomaly detection methods and observed the average area under the receiver operating characteristic curve increase of 3% on MVTec anomaly detection.

**Keywords:** generative adversarial networks; reconstruct; anomaly detection; sparsity

---

### 1. Introduction

Anomaly detection is widely used in various fields. In industry [1], it is employed to address safety and security concerns. For instance, it is also very important to ensure quality in the case of materials production, while in the medical field, it is utilized to analyze medical images, such as

identifying the location of tumors [2]. One of the primary challenges in anomaly detection is the scarcity of abnormal samples, which complicates the acquisition of representative and comprehensive data for training and testing purposes. Furthermore, anomalies tend to be unpredictable and diverse, making it challenging to capture all possible variations and features of abnormal regions. Additionally, the process of labeling and calibrating datasets is both costly and time-consuming. These factors suggest that traditional supervised learning methods are not well-suited for anomaly detection, and that unsupervised or weakly-supervised learning methods are more appropriate.

In recent years, deep learning techniques have been widely applied to various reconstruction tasks [3], leading to the development of numerous anomaly detection methods that rely on reconstruction quality [4]. Some methods [5] attempt to reconstruct abnormal samples as normal samples by learning the distribution of normal samples. Using a threshold method, the regions with large differences between the abnormal samples and the reconstructed samples are identified as defects. Other methods [6] use artificially generated defects for training, which weakens the reconstruction of anomalous regions. However, due to the difficulty of learning real defect features, these methods do not generalize well enough for real defects and can lead to overfitting.

To address these issues, in this paper, we propose a weakly-supervised anomaly detection method (WADS) that utilizes sparsity priority. We employ the Perlin noise function to generate random anomaly regions and construct various types of anomaly samples based on normal samples. In particular, we utilize sparsity prior as a regularization term to mitigate overfitting of the model and to enhance its generalization to real anomalies. In summary, we make the following contributions:

- **The new segmentation subnetwork** provides segmentation information for the generating subnetworks and addresses the overfitting issue prevalent in anomaly detection methods based on generative models.
- **A sparse regular term based on kullback-leibler (KL) divergence** is designed to constrain the area of the abnormal region, provide prior information about the abnormal region to the model, and enhance the model's generalization capabilities.

## 2. Related work

**Anomaly detection:** Many researchers work a lot for anomaly detection tasks. Traditional anomaly detection methods, such as principal component analysis and random projection [7], is to downscale the input data. The downscaled data contains main features, and the anomalous image will lose more information in the downscaling process leading to a larger reconstruction error, through which the anomaly detection is completed. With the development of deep learning, many deep learning methods in anomaly detection have gained better performance than traditional methods. Among them, methods based on image reconstruction work very well.

**Reconstruction Methods:** Methods based on reconstruction utilize the properties of reconstructed models, such as auto-encoder (AE) [8] and generative adversarial networks (GAN) [9]. By learning the distribution of positive samples, it can reconstruct negative samples into positive ones. Comparing the reconstructed samples with the negative samples, the regions with larger differences can be considered as anomalous regions. AE is widely used in image reconstruction. It will map the input image into a hidden space vector, which can represent the features of the image. Then the hidden space vector is decoded into a reconstructed image. GAN is similar to AE, but it uses an adversarial

loss function for training and calculates anomaly scores based on the image reconstruction effect for improving the reconstruction performance. Many methods combine the ideas of AE and GAN. GANomaly [10] improves on anoGAN by adding the structure of AE to the generator. Skip-GAN [11], in order to improve the reconstruction performance, further incorporates U-Net structure in the generator for feature fusion. However the reconstruction performance is so strong that anomalous regions are also reconstructed. TricycleGAN [12], proposed in 2021, implements a picture-to-picture image segmentation network, which accomplishes anomaly detection through three GANs. However, it needs to know the shape of the anomaly region, which means it can only segment for a certain class of shape.

Reconstruction by inpainting based anomaly detection method (RIAD) [13] proposes a novel anomaly detection method based on image inpainting. It reconstructs anomalous regions by utilizing information from normal regions, and the difference between the reconstructed image and the original image highlights the anomalous regions. Multi-resolution knowledge distillation (MKD) [14] proposes a novel anomaly detection method based on knowledge distillation. It distills knowledge from a teacher network trained on normal data to a student network at multiple resolutions, and anomalies are detected by the inconsistency between the student and teacher predictions. RevDistill [15] proposes a novel anomaly detection method based on reverse knowledge distillation. It first trains an embedding network on normal data, and then distills knowledge from the embedding network to a student network, which is trained to reconstruct normal data. Anomalies are detected by the reconstruction error of the student network. RealNet [16] proposes a framework for anomaly detection using pretrained models. The framework includes three key components: strength-controllable diffusion anomaly synthesis (SDAS), anomaly-aware features selection (AFS), and reconstruction residuals selection (RRS), which are used to generate synthetic anomaly samples, select key features, and screen reconstructed residuals, respectively. Through the synergy of these components, RealNet is able to achieve significant performance improvements in anomaly detection tasks.

**Segmentation methods:** In order to extract features from reconstructed and original images, reconstruction-based methods use image segmentation models for anomaly localization. Fully convolutional networks (FCN) [17], a landmark algorithm in image segmentation, reduces the computational effort by converting all fully connected layers of traditional convolutional neural networks (CNN) into convolutional layers. However, FCN uses the last layer of feature maps for prediction, and the segmentation for small targets is not good. Feature pyramid network (FPN) [18] utilizes the idea of image pyramid to construct multi-scale features and improve segmentation performance by processing different scale features. U-Net [19], which works well in the medical field, has a similar structure to FPN. It fuses multi-scale feature maps, and the overall structure is similar to a “U”, with the first half completing feature extraction and the second half completing upsampling.

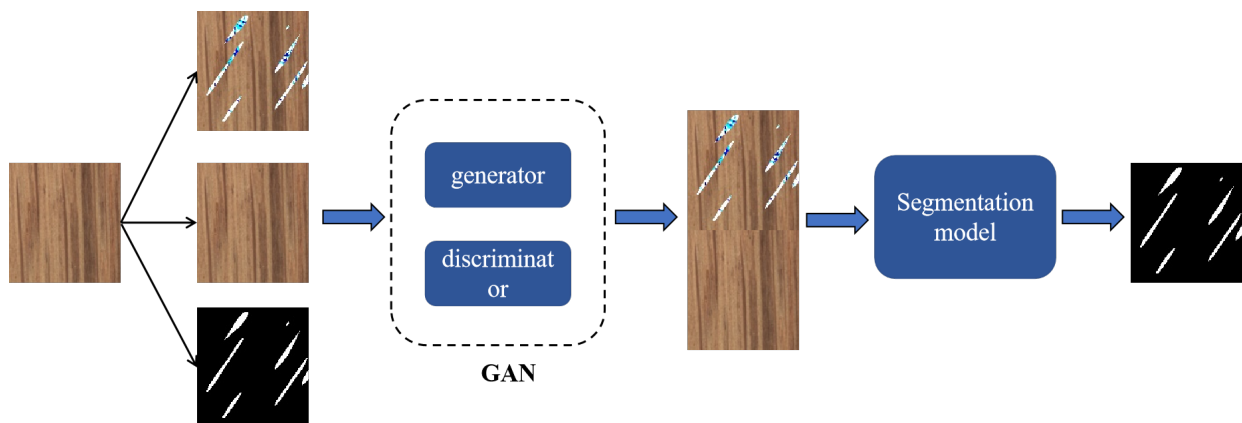
### 3. WADS

The proposed anomaly detection method (WADS) comprises a reconstructive subnetwork and a segmental subnetwork. The reconstructive subnetwork learns the distribution of normal regions and identifies regions that deviate from this normal distribution as abnormal. It then reconstructs these abnormal regions to resemble normal regions. The resulting reconstructed image serves as input for

the segmental subnetwork. By comparing the features of the reconstructed image with those of the original input image, the segmental subnetwork identifies the regions that significantly differ from the original, thereby completing the defect segmentation task.

### 3.1. The overview of WADS

The procedure of the proposed method is shown in Figure 1. The procedure is divided into three parts: 1) Processing normal samples to obtain abnormal samples and abnormal region masks; 2) using a GAN to reconstruct the input image and obtain the reconstructed image; 3) the segmental subnetwork processes both the input image and the reconstructed image to predict the mask of the anomalous region.



**Figure 1.** Method Procedure.

In particular, the reconstructive subnetwork and the segmental subnetwork should be trained together during the training process because they are interdependent. The reconstructive subnetwork needs to reconstruct the image to facilitate easier segmentation, while the segmental subnetwork needs to compare the features of the reconstructed image with those of the original image. During back propagation, their loss functions are combined and optimized concurrently.

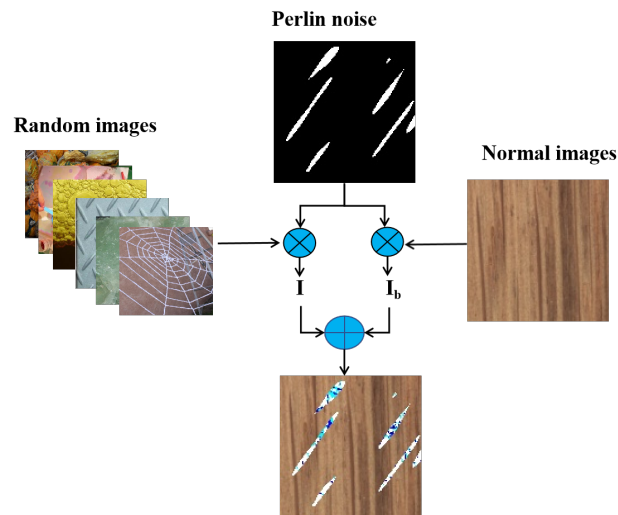
### 3.2. Dataset generation

The training dataset requires only normal images, as we generate abnormal regions that deviate from the normal distribution, which can be used to construct abnormal samples. Specifically, the final dataset comprises normal images, abnormal images, and abnormal area mask images.

Normal images and random images serve as the basis for constructing the dataset. The distribution of the random image is independent of the distribution of the normal image, so the area in the random image can be considered as an abnormal area in the abnormal image. All processes are depicted in Figure 2. First, due to the randomness of the abnormal area, Perlin [20] noise is utilized to construct the mask image of the abnormal area. Then, according to the mask image, the corresponding areas in the random image are extracted. Finally, image fusion is performed between the extracted area and the normal image to obtain the abnormal image  $I_a$ , which can be expressed as Eq (3.1):

$$I_a = \beta * I_b + (1 - \beta) * I \quad (3.1)$$

where  $I_b$  is the normal image and  $I$  is the random image after separation. In addition, we also performed rotation, scale transformation, and other operations on the normal images for dataset enhancement. According to processes, we can finally get normal pictures, abnormal pictures and abnormal region masks.



**Figure 2.** Dataset generation process.

### 3.3. Reconstructive subnetwork

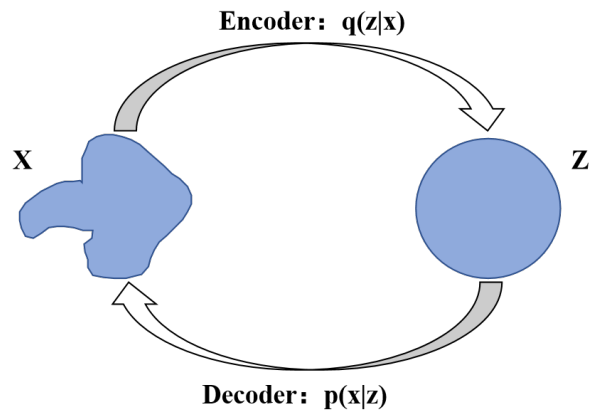
The reconstructive subnetwork mainly performs the task of image reconstruction in the whole approach, using the structure of GAN. In fact, variational auto-encoder (VAE) [21] also has generative properties, but VAE requires that the encoding of the hidden space needs to conform to a specific distribution, such as a mixed Gaussian distribution. The VAE structure is shown in Figure 3. To approximate the true distribution, VAE uses KL divergence to define the optimization objective, which is to maximize:

$$Loss = \int p(z|x) \log(q(x|z)) dz - \int p(z|x) \log\left(\frac{p(z|x)}{p(z)}\right) dz \quad (3.2)$$

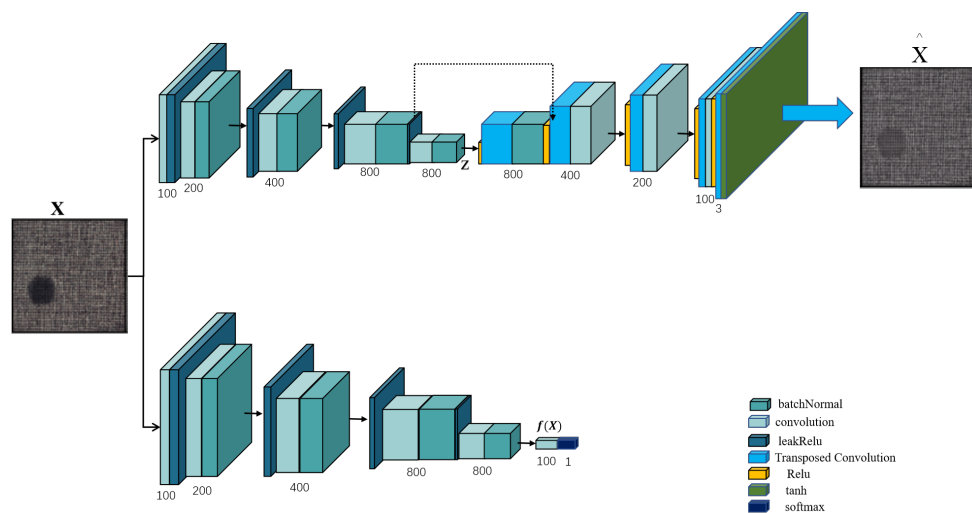
where  $p(z)$  is a given distribution. From Eq (3.2), VAE requires forcing the inputs data to be fitted to a specific distribution, which can create some problems: First, information will be lost in the mapping process. Second, input data that does not conform to the preset distribution will be poorly encoded and reconstructed. For reconstructed images, it will be blurred. GAN, on the other hand, does not have strong preconditioning like VAE; it directly fits the hidden space code to the known data distribution and uses discriminator network for optimization, so it can have better reconstruction for anomalous samples.

The structure of reconstructive subnetwork is mainly borrowed from skip-GAN, which adds encoder and decoder structure to the generator structure to enhance the reconstruction performance. The model structure is shown in Figure 4. The main body of the generator structure is the encoder and decoder, combined with the structure of feature fusion. The original structure of skip-GAN fuses the features of each layer of the encoder, so the reconstruction performance is too strong to reconstruct abnormal

regions into normal regions. We only input the last layer of the encoder feature into the decoder. The discriminator structure is relatively simple and completes the binary classification task.



**Figure 3.** VAE structure. Where  $X$  is inputs,  $Z$  is latent variable.



**Figure 4.** Reconstructive subnetwork.

The loss function of the generator consists of three parts:

$$\begin{aligned}
 L_{con} &= SSIM(X, \hat{X}) + L_2(X, \hat{X}), \\
 L_{lat} &= L_2(Z, \hat{Z}), \\
 L_{adv} &= BCE(f(X), f(\hat{X}))
 \end{aligned}
 \tag{3.3}$$

where  $X$  is the input image,  $\hat{X}$  is the reconstructed image,  $Z$  is the output of the normal image through the encoder,  $\hat{Z}$  is the output of the abnormal image through the encoder,  $f(X)$  is the feature output of the normal image at the last layer of the discriminator, and  $f(\hat{X})$  is the feature output of the reconstructed image at the last layer of the discriminator.  $L_{con}$  and  $L_{lat}$  guarantee the reconstruction performance of the generator.  $L_2$  loss, which reflects the overall difference between the reconstructed image and

the original image, is often used in reconstruction methods. However, it considers that each pixel is independent, without considering the relationship between pixels. Therefore, we add structural similarity index measure (SSIM) loss [22], which calculates the loss based on patches, taking the connection between pixels into account. The reconstruction performance is guaranteed while limiting the reconstruction of abnormal regions in terms of luminance, contrast and structure. The final loss function of the generator is:

$$L_{total} = \lambda_{adv} * L_{adv} + \lambda_{con} * L_{con} + \lambda_{lat} * L_{lat}, \quad (3.4)$$

where  $\lambda_{adv}, \lambda_{con}, \lambda_{lat}$  are the weights of the three loss functions, respectively.

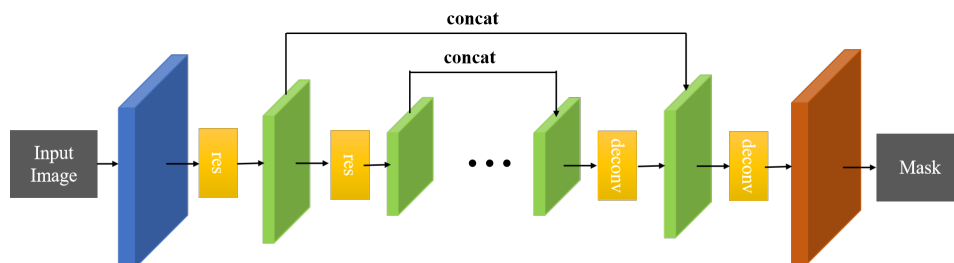
The discriminator is mainly used for binary classification. It uses binary cross entropy (BCE) loss, which is:

$$L_{disc} = BCE(y, \hat{y}), \quad (3.5)$$

where  $y$  is the real label of the input image and  $\hat{y}$  is the output label of the discriminator. The discriminator will determine whether the input is a reconstructed image.

### 3.4. Segmental subnetwork

The segmental subnetwork performs the anomaly segmentation task, outputting a masked map of the anomaly region. The model structure is shown in Figure 5. It draws on the structure of FPN [18]. First, residual network 50 (ResNet50) [23] with strong feature extraction ability is used to downsample the input, followed by upsampling the downsampled image. Finally, after sigmoid activation, the output is the mask of the anomalous region. In the upsampling process, features of downsampling layers are used for feature fusion. By comparing the original and reconstructed image, the areas with large differences are considered as abnormal regions.



**Figure 5.** Segmentation model structure.

The segmental subnetwork actually performs a binary classification task for each pixel in the image. BCE loss is often used in the binary classification. However, in anomaly detection task, the abnormal region accounts for a small proportion of the whole image, so the number of abnormal points are smaller than the normal points, resulting in an unbalanced problem. Using BCE loss will get bad results. Focal loss [24], which can increase the weights of a few categories in training, perfectly solves the problem of imbalance. In addition, for the sparsity defect, we design a special sparsity loss based on KL divergence [25]. KL divergence describes the distance between the two distributions. Sparsity loss is as follows:

$$L_{kl} = M * \log\left(\frac{M}{\rho}\right) + (1 - M) * \log\left(\frac{1 - M}{1 - \rho}\right), \quad (3.6)$$

where  $M$  is the abnormal area mask image and  $\rho$  is a constant, representing the proportion of abnormal areas, which limits the area of abnormal areas in the mask image. The total loss function is:

$$FL(y, \hat{y}) = \begin{cases} -(1 - \hat{y})^\rho \log \hat{y}, & y = 1 \\ -\hat{y}^\rho \log(1 - \hat{y}), & y = 0 \end{cases}, \quad (3.7)$$

$$L_{seg} = FL(y, \hat{y}) + L_{kl}$$

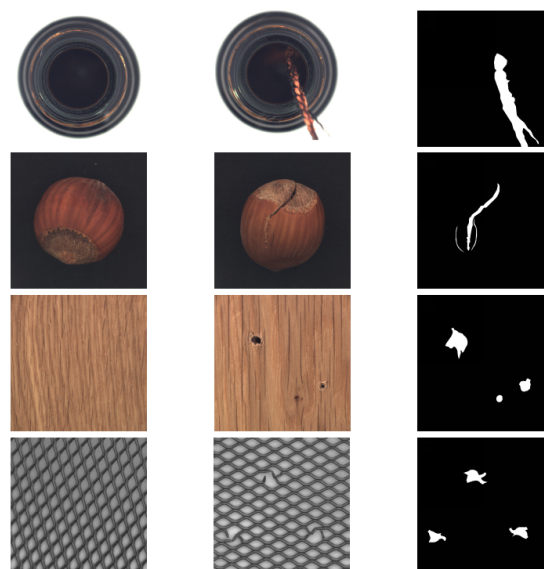
where  $y$  is the ground truth of segmentation and  $\hat{y}$  is the output of the segmentation model.

## 4. Experiments

The proposed method is widely evaluated and compared with other generative anomaly segmentation methods. The effects of SSIM and KL divergence loss function were verified by an ablation study.

### 4.1. Dataset and setup

Our proposed method is evaluated on MVTEC anomaly detection (MVTECAD) [26], which is often used for the evaluation of weakly supervised or unsupervised anomaly detection algorithms. The dataset contains 15 kinds of data, and some of which are shown in Figure 6. The training set only contains normal images, and the testing set contains different types of defects that gives a comprehensive picture of the capabilities and bounds of our method. In order to compare with others, the final dataset uses  $128 \times 128$  size images, with random cropping, scale transformation, and random rotation strategies for large size images, resulting in a final data of around 20000 per kind.



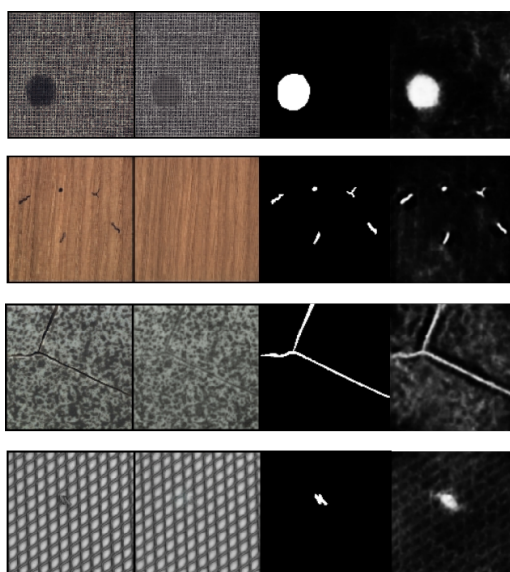
**Figure 6.** Examples of MVTEC. Normal images (left), abnormal images (center) and mask images (right) are shown in the figure.



## 4.2. Results

During the training process, the hyperparameters in the algorithm are set uniformly. To make the loss functions have the same order of magnitude,  $\lambda_{adv}$  is set to 1,  $\lambda_{con}$  is set to 20,  $\lambda_{lat}$  is set to 1,  $\rho$  is set to 0.1, the learning rate is set to 0.0001, and the encoding length of the hidden space of the encoder is set to 200. The model is trained for 700 epochs. For evaluation, we choose the pixel-based area under the receiver operating characteristic curve (AUROC) to evaluate the performance of our method. The larger the AUROC value, the better the model performance.

Figure 7 shows some experimental results of WADS. As can be seen from the figure, the generation model can reconstruct the normal regions well, and the less obvious anomalous regions that can be removed, and the more obvious ones can be weakened significantly. The segmentation model can also get an accurate segmentation image of abnormal regions.



**Figure 7.** Outputs of the method. Input images (first), reconstructed images (second), ground truth images (third) and, output mask images (fourth) are shown in the figure.

In Tables 1–3, the comparison results of our method and other methods are shown. From the table, we can see that our method achieves better results in all surface anomaly detection, such as wood, leather, etc. In object anomaly detection, the results are not uniform. It can achieve particularly good results on toothbrush and hazelnut, but particularly poor results on zipper. This occurs because the SSIM loss function is concerned with human vision, which drives the generation model to reconstruct the image to visually approximate a normal image. However, on the zipper dataset, the anomalous regions are so weak that it is difficult for human vision to detect the anomalous regions, resulting in the reconstructed images not being well cleared of the anomalous regions. Other methods do not achieve segmentation by comparing the differences between the reconstructed images and the input images, so the above problems do not exist. On the visually distinguishable anomalous dataset, our method achieves better results. Therefore, the segmentation model gets bad results. Other methods focus on distribution, so the above problems do not exist. On visually distinguishable abnormal datasets, our method has achieved better results.

**Table 1.** AUROC in % in MVTecAD.

Class	Ours	Ustd [27]	$\gamma$ -VAE grad [28]	RAID [13]	MKD [14]	P-Net [29]	GDR [28]	RevDistill [15]
<b>Mean</b>	93.0	85.7	88.8	91.7	90.7	88.7	89.5	91.4

**Table 2.** AUROC in % on object images in MVTecAD.

Class	Ours	Ustd	$\gamma$ -VAE grad	RAID	MKD	P-Net	GDR	RevDistill
bottle	97.0	91.8	93.1	99.9	99.4	99.0	92.0	46.2
capsule	80.1	91.6	91.7	88.4	80.5	84.0	92.0	97.3
pill	94.6	93.5	93.5	83.8	82.7	91.0	93.0	94.9
transistor	94.6	70.1	93.1	90.9	85.6	82.0	92.0	91.5
zipper	73.8	93.3	87.1	98.1	93.2	90.0	87.0	98.9
cable	95.6	86.5	88.0	81.9	89.2	70.0	91.0	76.9
hazelnut	99.2	93.7	98.8	83.3	98.4	97.0	98.0	100.0
metal nut	93.9	89.5	91.4	88.5	73.6	79.0	91.0	94.9
screw	93.6	92.8	97.2	84.5	83.3	99.9	95.0	96.5
toothbrush	99.1	86.3	98.3	100.0	92.2	99.0	99.0	83.6
<b>Mean</b>	92.2	88.9	93.2	89.9	87.8	89.1	93.0	88.1

**Table 3.** AUROC in % on texture images in MVTecAD.

Class	Ours	Ustd	$\gamma$ -VAE grad	RAID	MKD	P-Net	GDR	RevDistill
grid	94.3	81.9	97.9	99.6	78.0	78.0	96.0	97.9
leather	96.4	81.9	89.7	100.0	95.1	95.1	93.0	100.0
tile	95.6	91.2	58.1	98.7	91.6	91.6	65.0	97.7
carpet	91.5	69.5	72.7	84.2	79.3	79.3	74.0	95.7
wood	94.5	72.5	80.9	93.0	94.3	94.3	84.0	98.9
<b>Mean</b>	94.5	79.4	79.9	95.1	87.7	87.7	82.4	97.9

### 4.3. Ablation study

To verify the effectiveness of SSIM and KL divergence loss functions, we conducted ablation experiments, respectively. The effect on model performance was observed by deleting SSIM in the loss function of the generator and deleting KL divergence loss function in the segmentation model. AUROC is used for performance evaluation, and results are shown in Table 4.

From the results, it can be seen that after removing the SSIM loss function, the model works poorly for some object anomaly detection, such as capsule, transistor, and screw. That is because that the reconstruction performance of the model is reduced, resulting in the inability to accurately determine the anomaly region. The effect is still good for surface anomaly detection, which is because the model can distinguish the anomalous region from the perspective of distribution. Specifically for zipper, without using SSIM, the effect of anomaly detection is better. After deleting the KL divergence loss function, the proportion of sparsity samples in each class testing set is different, so the results are not

the same. For the bottle with more sparsity samples, after deleting the KL divergence loss function, the AUROC decreases by about 10%, but for the grid and tile with less sparsity samples, the AUROC increases. This is in line with our expectations. From the ablation study, it can be seen that the SSIM loss function does guide the reconstruction of the model from the visual angle, and weakens the influence of the distribution on the reconstruction. KL divergence loss function does have better effect for sparsity abnormal samples.

To verify the effectiveness of the discriminant subnetwork, we also conducted ablation experiments on it. The ablation experiment in this section primarily compares the changes in structured sparse bayesian gaussian (SSBG) AUROC on MVTecAD after post-processing with the discriminant subnetwork and the traditional local binary patterns (LBP) algorithm. The experimental results are presented in Table 4.

**Table 4.** Ablation Results for SSIM, KL, and subnetwork on the MVTec dataset (AUROC).

Class	Method	Method <sub>noSSIM</sub>	Method <sub>noKL</sub>	Method <sub>LBP</sub>
bottle	97.0	82.2	87.7	94.3
capsule	80.1	51.6	78.6	73.6
grid	94.3	93.5	97.2	85.6
leather	96.4	99.1	94.2	93.7
pill	94.6	92.8	86.8	84.8
tile	95.6	98.7	97.1	92.3
transistor	94.6	67.7	89.2	90.2
zipper	73.8	84.8	78.5	59.3
cable	95.6	96.2	94.1	87.9
carpet	91.5	88.9	90.2	86.4
hazelnut	99.2	93.4	97.9	95.8
metal nut	93.9	74.9	82.3	87.9
screw	93.6	59.9	96.3	86.8
toothbrush	99.1	98.4	96.4	96.4
wood	94.5	91.5	90.6	90.7
<b>Mean</b>	93.0	84.9	90.7	87.0

As seen in the table, the post-processing effect of the traditional LBP algorithm is inferior to that of the discriminant subnetwork, especially in categories where the target area is not distinct, such as capsules and zippers, where the traditional LBP algorithm performs poorly. This is because there is a discrepancy between the reconstructed background region and the real background region, and the traditional algorithm's feature extraction capability is weaker compared to the discriminant subnetwork. Consequently, when the difference between the target region and the background region is minimal, the traditional algorithm's segmentation performance is significantly less effective than that of the deep learning method. Overall, since the discriminant subnetwork has superior feature extraction capabilities and can better measure the distance between the target region and the background region, using the discriminant subnetwork for segmentation is more efficient than using the traditional algorithm.

## 5. Conclusions

In this paper, we propose a WADS based on sparsity-constrained reconstruction. To enhance the reconstruction quality of the reconstructive subnetwork, we employ a hybrid of GAN and AE. We also adopt SSIM loss function to quantify the discrepancy between the reconstructed image and the original image. To tackle the over-fitting issue of the generative model-based anomaly detection method, WADS utilizes a segmental subnetwork to supply segmentation information for the reconstructive subnetwork. Specifically, we devise a sparsity regularized based on KL divergence, which constrains the size of anomaly regions and prioritizes anomaly regions for the model, to improve the generalization of WADS. Overall, our method achieves a superior segmentation performance for visually salient anomalies, especially for sparsity defects. However, since SSIM loss mimics human perception by primarily focusing on edge and texture similarity, its detection performance is insufficient for subtle and visually unobtrusive anomalies. In the future, we will further investigate the problem of detecting weak anomalous regions.

### Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

This research was supported by NSFC (No. 61871074).

### Conflict of interest

The authors declare that there are no conflicts of interest.

## References

1. W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, K. Müller, Explaining deep neural networks and beyond: A review of methods and applications, in *Proceedings of the IEEE*, **109** (2021), 247–278. <https://doi.org/10.1109/JPROC.2021.3060483>
2. G. Quellec, M. Lamard, M. Cozic, G. Coatrieux, G. Cazuguel, Multiple-instance learning for anomaly detection in digital mammography, *IEEE Trans. Med. Imaging*, **35** (2016), 1604–1614. <https://doi.org/10.1109/TMI.2016.2521442>
3. S. Zhang, J. Li, L. Yang, Survey on low-level controllable image synthesis with deep learning, *Electron. Res. Arch.*, **31** (2023), 7385–7426. <https://doi.org/10.3934/era.2023374>
4. S. Venkataramanan, K. Peng, R. V. Singh, A. Mahalanobis, Attention guided anomaly localization in images, in *Computer Vision – ECCV 2020*, **12362** (2020), 485–503. [https://doi.org/10.1007/978-3-030-58520-4\\_29](https://doi.org/10.1007/978-3-030-58520-4_29)

5. T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, U. Schmidt-Erfurth, f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks, *Med. Image Anal.*, **54** (2019), 30–44, <https://doi.org/10.1016/j.media.2019.01.010>
6. L. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, preprint, arXiv:1706.05587.
7. E. Bingham, H. Mannila, Random projection in dimensionality reduction: Applications to image and text data, in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, (2001), 245–250, <https://doi.org/10.1145/502512.502546>
8. T. Tang, W. Kuo, J. Lan, C. Ding, H. Hsu, H. Young, Anomaly detection neural network with dual auto-encoders gan and its industrial inspection applications, *Sensors*, **20** (2022), 3336. <https://doi.org/10.3390/s20123336>
9. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, preprint, arXiv:1406.2661.
10. S. Akçay, A. Atapour-Abarghouei, T. P. Breckon, GANomaly: Semi-supervised anomaly detection via adversarial training, in *Computer Vision-ACCV 2018*, **11363** (2018), 622–637. [https://doi.org/10.1007/978-3-030-20893-6\\_39](https://doi.org/10.1007/978-3-030-20893-6_39)
11. S. Akçay, A. Atapour-Abarghouei, T. P. Breckon, Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection, in *2019 International Joint Conference on Neural Networks (IJCNN)*, (2019), 1–8. <https://doi.org/10.1109/IJCNN.2019.8851808>
12. U. Sivanesan, L. H. Braga, R. R. Sonnadara, K. Dhindsa, TricycleGAN: Unsupervised image synthesis and segmentation based on shape priors, preprint, arXiv:2102.02690.
13. V. Zavrtnik, M. Kristan, D. Skočaj, Reconstruction by inpainting for visual anomaly detection, *Pattern Recognit.*, **112** (2021), 107706. <https://doi.org/10.1016/j.patcog.2020.107706>
14. M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, H. R. Rabiee, Multiresolution knowledge distillation for anomaly detection, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 14897–14907. <https://doi.org/10.1109/CVPR46437.2021.01466>
15. H. Deng, X. Li, Anomaly detection via reverse distillation from one-class embedding, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 9727–9736. <https://doi.org/10.1109/CVPR52688.2022.00951>
16. X. Zhang, M. Xu, X. Zhou, RealNet: A feature selection network with realistic synthetic anomaly for anomaly detection, preprint, arXiv:2403.05897.
17. E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
18. T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 936–944. <https://doi.org/10.1109/CVPR.2017.106>
19. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*, **9351** (2015), 234–241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

20. K. Perlin, An image synthesizer, *ACM SIGGRAPH Comput. Graphics*, **19** (1985), 287–296, <https://doi.org/10.1145/325165.325247>
21. D. P. Kingma, M. Welling, Auto-encoding variational bayes, preprint, arXiv:1312.6114.
22. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, **13** (2004), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
23. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
24. T. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>
25. O. Cappé, A. Garivier, O. Maillard, R. Munos, G. Stoltz, Kullback-Leibler upper confidence bounds for optimal sequential allocation, *Ann. Stat.*, **41** (2013), 1516–1541, <http://doi.org/10.1214/13-AOS1119>
26. P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, MVTEC AD-A comprehensive real-world dataset for unsupervised anomaly detection, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 9584–9592. <https://doi.org/10.1109/CVPR.2019.00982>
27. P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 4182–4191. <https://doi.org/10.1109/CVPR42600.2020.00424>
28. D. Dehaene, O. Frigo, S. Combrexelle, P. Eline, Iterative energy-based projection on a normal data manifold for anomaly localization, preprint, arXiv:2002.03734.
29. K. Zhou, Y. Xiao, J. Yang, J. Cheng, W. Liu, W. Luo, et al., Encoding structure-texture relation with P-Net for anomaly detection in retinal images, in *Computer Vision-ECCV 2020*, **12365** (2020), 360–377. [https://doi.org/10.1007/978-3-030-58565-5\\_22](https://doi.org/10.1007/978-3-030-58565-5_22)



AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)