IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 November 2023, accepted 1 December 2023, date of publication 5 December 2023,
date of current version 13 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339780

RESEARCH ARTICLE

SCGAN: Extract Features From Normal Semantics for Unsupervised Anomaly Detection

YANG DAI[1], LIN ZHANG[2], FU-YOU FAN[3,4], YA-JUAN WU[1], AND ZE-KUAN ZHAO[1]
[1]School of Computer Science, China West Normal University, Nanchong 637000, China
[2]Sichuan Digital Economy Research Institute, Yibin 644000, China
[3]Network and Library Information Center, Yibin University, Yibin 644000, China
[4]Key Laboratory of Intelligent Terminal in Sichuan Province, Yibin 644000, China

Corresponding author: Ya-Juan Wu (scwuyajuan@163.com)

This work was supported in part by the Research Foundation for Talents of China West Normal University under Grant 17YC163, in part by the Open Project of Key Laboratory of Intelligent Terminal of Sichuan Province under Grant SCITLAB-0019, in part by the Open Project of Key Laboratory of Network and Data Security of Sichuan Province under Grant NDSZD201603, and in part by the Key Project of Education Department of Sichuan Province under Grant 17ZA0452.

ABSTRACT Anomaly detection within the realm of industrial products seeks to identify regions of image semantics that deviate from established normal patterns. Given the inherent challenges associated with collecting anomaly samples, we exclusively extract features from normal semantics. Our proposed solution involves a Semantic CopyPaste based Generative Adversarial Network (SCGAN) for unsupervised anomaly detection. To enable the comprehensive acquisition of semantic features within intricate real-world images, we embrace an encoder-decoder-encoder as the fundamental network structure. In practical terms, our approach commences with the input image being subjected to the CopyPaste augmentation module. Here, we strategically copy N patches, each constituting 1% of the image's area, from normal samples. These patches are then randomly pasted into different regions of the original image. Subsequently, a generative adversarial network is trained to facilitate sample reconstruction. A noteworthy augmentation to the network's channel attention capabilities entails the incorporation of a multi-scale channel attention module within the first encoder. This module serves to emphasize contextual features across varying scales within the image. During the test, we detect anomalous regions by meticulously comparing residuals between the input image and its reconstructed counterpart. Our methodology is rigorously validated through diverse experiments conducted on challenging MVTec and BTAD public datasets. The results conclusively affirm the state-of-the-art performance achieved by our proposed method in the domain of anomaly detection.

INDEX TERMS Anomaly detection, generative adversarial networks, channel attention, multi-scale channel attention module.

I. INTRODUCTION

Industrial image anomaly detection is a widely used technique for identifying product defects throughout the manufacturing and production processes. These methods facilitate real-time identification and localization of surface anomalies in products. By effectively filtering out anomalies, they decrease the incidence of defective products. Furthermore, in daily operations, anomaly detection can play a pivotal role in forecasting equipment failures. This, subsequently, reduces maintenance costs by issuing timely alerts to enterprises, enabling them to take proactive preventive measures.

In the realm of manufacturing, an anomaly refers to the presence of abnormal semantic pixels within a product's surface image, which could manifest in any part of the image. Anomaly detection [1] is the process of identifying image data that significantly deviates from typical instances. Currently, the scope of anomaly detection encompasses a variety of applications, including defect detection [2], medical diagnosis [3], [4], video surveillance [5], [6], financial transaction monitoring [7], and network security [8]. Of particular note, anomaly detection finds extensive utility within the industrial

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

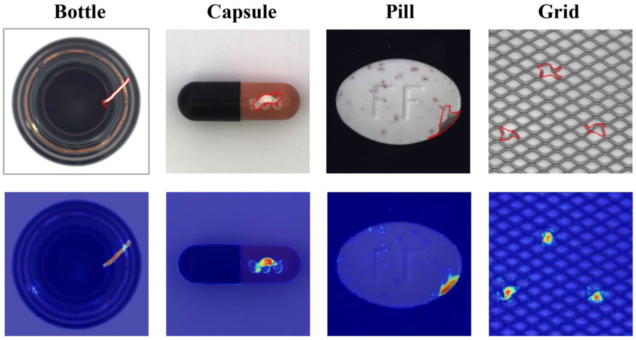


FIGURE 1. The first row is the ground truth, where the red boundary indicates the anomalous contours in the real world. The second row is the heat map of the segmentation result of SCGAN.

sector, spanning applications such as detecting conductive particles in glass chips [9] and inspecting steel surfaces [10].

Anomalies can stem from various factors, including defects in raw materials, equipment malfunctions, or process immaturity. Within the manufacturing industry, anomaly detection serves as a crucial tool for product monitoring. Its operation has two functions: firstly, by effectively averting defective products from entering the market, thereby elevating product quality; secondly, by providing enterprises with timely insights through test results. Consequently, anomaly detection emerges as a pivotal step within industrial production. There are currently four important challenges in this detection task [11]: i) The data of various samples is unbalanced. ii) Ambiguities in defining decision boundaries. iii) Abnormal metric. iv) Acquiring knowledge pertaining to out-of-distribution (OOD) anomalies.

Our proposal introduces an unsupervised anomaly detection method grounded in semantic CopyPaste. This method only extracts features from normal semantics, with a primary focus on the semantic links and constraints existing among image pixels. Our assertion is that networks exhibiting poor image recovery capabilities should theoretically be considered indicative of anomalies.

With this objective in mind, we introduce an unsupervised anomaly detection network named SCGAN, founded on generative adversarial networks. SCGAN operates by executing reconstruction through the acquisition of feature semantics from normal samples. Notably, the network places a greater emphasis on discerning the intrinsic semantic links interconnecting object contexts. The outcomes of our method are vividly depicted in Figure 1, showcasing detection results on a public dataset. As observed, our method's detection outcomes closely align with the ground truth. By attuning itself to the semantics of contextual features, SCGAN aptly identifies and pinpoints anomalies within industrial settings. In the operational procedure, a normal sample serves as input, initiating its journey through the CopyPaste image enhancement module. This step deliberately introduces anomalies, with the anomaly's appearance randomized to simulate its possible occurrence anywhere on the product's surface. The augmented image subsequently traverses through a generative adversarial network for image reconstruction. Notably, we conducted experiments encompassing four distinct artificial anomaly enhancement scenarios, all aimed at nudging the network to grasp the normal semantics inherent in the image.

In summary, our contributions can be outlined as follows:

- We propose a novel unsupervised anomaly detection network called SCGAN, specifically designed to extract sample features from normal semantics only.
- We have developed an image enhancement module named CopyPaste, which serves the dual purpose of simulating anomalies in an artificial yet representative manner and assisting in semantic reconstruction for network.
- The experimental results demonstrate the remarkable performance of our method on two prominent public datasets, namely MVTec and BTAD.

II. RELATED WORKS

Deep learning-based approaches have gained significant traction in the field of anomaly detection, encompassing techniques like auto-encoders (AE) [12], convolutional neural networks (CNN) [13], and generative adversarial networks (GAN) [14]. Given the widespread adoption of GAN for anomaly detection [3], a plethora of network models characterized by exceptional performance have emerged over recent years.

Unsupervised deep anomaly detection entails the identification of anomalies solely based on the feature semantics derived from normal samples. Unsupervised methods hold an inherent advantage in this domain due to the challenging nature of acquiring labeled anomaly data. In recent years, the widespread adoption of unsupervised methods has established their superiority over traditional approaches. For instance, CBiGAN [15] represents an improvement upon BiGAN [16], addressing issues related to sample reconstruction. Notably, this enhancement introduces consistency constraints within the encoder and decoder of BiGAN, thus enhancing the modeling capability and precision of the reconstruction process. Yan et al. [17] contribute with a semantic context-based anomaly detection network named SCADN. This distinctive network incorporates random strip masks of varying widths and orientations into the input image, serving to encompass a broader context. Subsequently, network reconstruction is conducted. Zhang et al. [18] introduce DeSTSeg, a network architecture hinged on the student-teacher framework. This innovation amalgamates the teacher network, student network, and segmentation network, yielding optimal average accuracy across multiple levels - from image to pixel to instance. Zaheer et al. [19] present OGNet, a two-stage anomaly detection framework. This framework constructs a network based on an encoder-decoder paradigm, effectively transforming the anomaly detection task into a pursuit of both low and high-quality sample reconstruction. DAGAN [20] employs a fusion of skip connections and dual autoencoders to successfully achieve industrial detection.

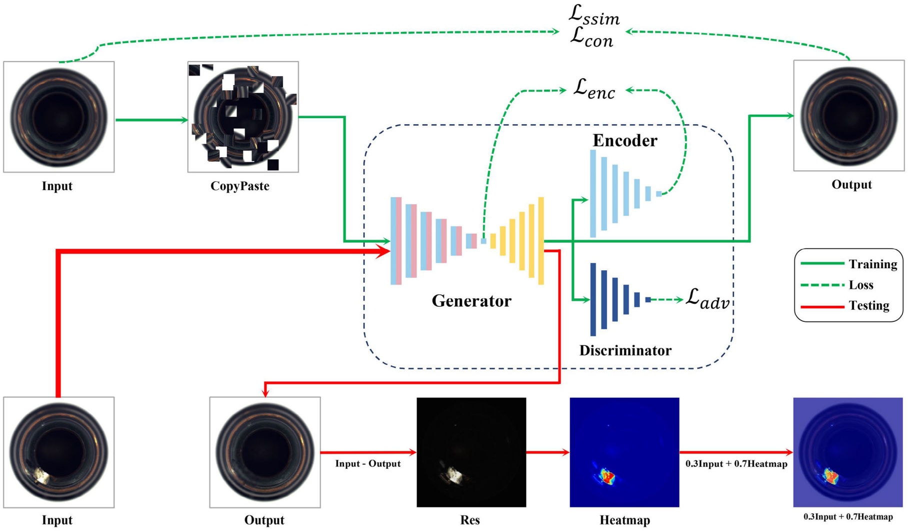


FIGURE 2. Network architecture of SCGAN. The training flow is shown as the green solid line. The network aims to randomly cut N masks with an area ratio of 1% from the input samples and then randomly paste them into the original image to obtain a CopyPaste image. The SCGAN is then trained to restore the image to its normal semantics. The flow of the testing is shown in the red solid line. The network uses the residuals between the input and output images to determine anomalies.

This strategic combination amplifies the network's capabilities in handling complex scenarios.

In parallel, self-supervised learning techniques find frequent application within detection tasks. Self-supervised learning capitalizes on extracting context-supervised information from the data itself, obviating the need for explicit labeling. Amid the array of self-supervised methods, the innovative approach of RIAD [21] transforms anomaly detection into a restoration reconstruction challenge for images. It achieves this by systematically removing content from specific regions within a partial grid and subsequently restoring the missing content based on the surrounding context. Furthermore, this study introduces a gradient similarity-based metric coupled with the loss strategy MSGMS. The efficacy of CutPaste [22] as a high-performance defect detection model hinges on its skillful data enhancement strategy. This strategy involves the random cropping of rectangular images of varying dimensions, which are then artfully inserted into different positions within the original image. Pirnay et al. [23] reframe anomaly detection as a patch-inpainting issue, proposing a method that leverages discarding convolutions and a self-attentive approach called InTra for effective reconstruction. The central concept revolves around repairing patches concealed by the network through the amalgamation of additional image information within a broader context. Ye et al. [24] introduce the attribute recovery framework ARNet, reshaping anomaly detection into an image recovery task. ARNet adeptly extracts semantic features by selectively erasing sample attributes like color and orientation during the training.

In the semi-supervised learning -based approach, GANomaly [25] introduced an encoder-decoder-encoder sub-network paradigm. This design involved mapping the image into low-dimensional vectors, followed by reconstruction to generate the final image, ultimately leading to mapping the generated image into the potential representation. Nevertheless, GANomaly's performance falls short in terms of adequately reconstructing intricate image details, posing challenges in handling realistic high-dimensional complex images. In a different vein, Mishra et al. [26] proposed a transformer-based image anomaly detection and localization network named VT-ADL. This innovative approach employs a Gaussian mixture density network to accurately pinpoint anomalies subsequent to the encoder's output.

One class of novelty detection also belongs to anomaly detection. An example of this is OCGAN [27], which is employed to ascertain whether a given sample originates from the same class as the training samples. The underlying principle of this model posits that in-class samples can be aptly represented, whereas out-of-class samples exhibit inferior representations.

III. PROPOSED METHOD

The method we propose hinges upon an encoder-decoder-encoder network, meticulously trained exclusively on normal samples for the extraction of normal semantic features. In the following section, we provide an initial introduction to the overarching model we have devised. Subsequently, we delve into an in-depth exploration of the constituent sub-modules within the model. This detailed breakdown predominantly encompasses the multi-scale channel attention module and the CopyPaste.

A. SCGAN

The structure of SCGAN is shown in Fig. 2. Throughout the training, the network exclusively processes normal samples. Subsequently, these samples are subjected to enhancement via the CopyPaste technique before being fed into the SCGAN. Finally the network undergoes adversarial reconstruction to get the output samples. Within the SCGAN, we adopt the foundational encoder-decoder-encoder. Upon completion of the comprehensive training regimen, the network acquires a trio of pivotal capabilities. First, it attains proficiency in mapping the original image into the latent space. Second, it gains the ability to translate latent vectors back into the image space. Lastly, the network develops the acumen to discern between normal and abnormal images.

In the testing, samples containing anomalies are input into the trained generator, yielding reconstructed outputs. Given that the network exclusively learns semantic features from normal samples, its reconstruction ability for abnormal pixels is limited. Consequently, a notable discrepancy emerges between the pre-reconstruction and post-reconstruction states, manifested as a substantial residual difference. This residual image is derived from the difference between the input and output images. Subsequently, a heat map is derived from this residual image. Moreover, the corresponding anomaly thermogram is generated utilizing this residual image. This involves overlaying the 0.7x heat map onto the 0.3x input image to facilitate observation. The generation of the heat map is contingent upon the probability matrix generated by the network model upon analyzing the test image. Different probabilities correspond to diverse heat colors within the map, spanning a gradient from blue to red. The extent of abnormality is proportionally reflected by the degree of red hue, with a stronger deviation from normalcy resulting in a more pronounced red coloring of the pixel region.

Figure 3 illustrates the detailed structure of the SCGAN. The generator's architecture commences with seven convolutional blocks, which extract feature maps via down-sampling. Subsequently, seven transposed convolutional blocks are employed to restore the feature maps to their original dimensions through up-sampling. Of particular significance is the incorporation of a multi-scale channel attention module within the down-sampling convolutional blocks. This module is strategically introduced to enable an attentional fusion of features across varying scales. The structure of the appended encoder aligns with that of the discriminator, except for the final layer. This additional decoder plays a pivotal role in diminishing the gap between the bottleneck features of both components, thereby significantly influencing the quality of image reconstruction.

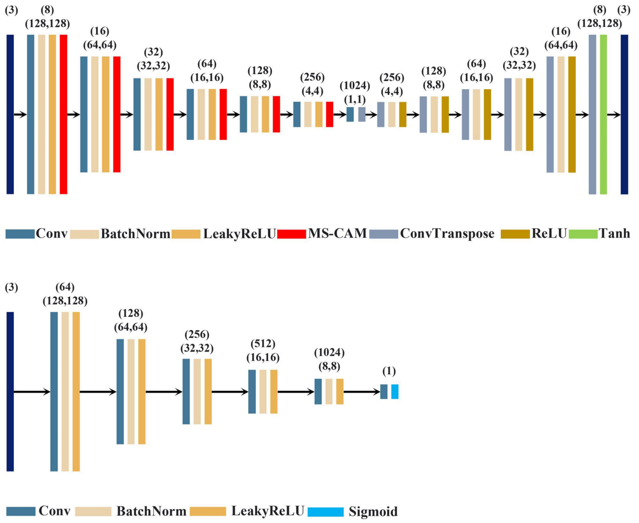


FIGURE 3. Detailed network structure of SCGAN, including network composition, feature map size and channel information.

B. MULTI-SCALE CHANNEL ATTENTION MODULE

The attention module has demonstrated its unique advantages in a wide range of deep learning studies. We have applied the attention mechanism to the industrial image anomaly detection task in order to further improve the model's extraction of feature semantics, inspired by Dai et al. [28]. The attention module can correlate global and local information and plays a key role in the transmission of image. The network uses the attention module to bring the distance between the pixels of training image and the generated image closer. The image is reconstructed based on the horizontal and vertical correlation between the pixels and then the residuals of both are used to determine whether it is anomalous or not.

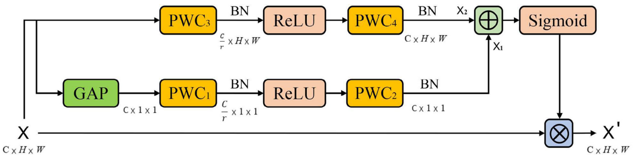


FIGURE 4. Structure of the multi-scale channel attention module.

In a specific, we introduce a Multi-Scale Channel Attention Module (MS-CAM) to every layer within the first encoder. This integrative module systematically extracts channel attention from the image across various scales. Distinguished by its capacity to establish meaningful relationships between global and local information, the attention module adeptly orchestrates suitable dependencies.

As shown in Fig. 4, MS-CAM focuses on the scaling problem of the channel through point-wise convolution, which is generally divided into global and local features. As far as global features are concerned, an input feature map X of size $H \times W$ with C channels undergoes initial processing. After global average pooling is applied to reduce feature dimensions, a feature map of size $C \times 1 \times 1$ is obtained. Subsequently, a point-wise convolution (PWC_1) with a convolution kernel size of 1×1 is employed for channel reduction, effectively decreasing the number of channels in the feature map to $1/r$ of the original count. Following the Batch Normalization (BN) layer and activation function, another point-wise convolution, PWC_2, is employed for channel recovery, restoring the number of channels to the original C. The result, X_1, is obtained after the BN layer.

Here, r represents the channel reduction ratio. We set r to 4 in our experiments. The channel attention calculation formula for its global feature is shown in equation (1):

$$X_1 = BN(PWC_2(ReLU(PWC_1(GAP(X))))) \tag{1}$$

where GAP denotes Global Average Pooling, $\mathrm{PWC}_{1,2}$ denotes point-wise convolution of different sizes, ReLU denotes activation function, and BN denotes batch normalization.

The process is similar to the previous one as far as local features are concerned, but the global average pooling operation is eliminated. The resulting feature map sizes after two point-wise convolutions are $(C/\mathrm{r})\times H \times W$ and $C \times H \times W$, respectively. The formula for calculating the channel attention of its local features is shown in equation (2):

$$X_2 = BN(PWC_4(ReLU(PWC_3(X)))) \tag{2}$$

Finally, the result of attention calculation of global and local features is broadcasted addition and sigmoid activation function, and then fused features with the original image X to get X'. Its fusion calculation formula is shown in equation (3):

$$X' = X \otimes (Sigmoid(X_1 \oplus X_2)) \tag{3}$$

Where $\otimes$ denotes the element-wise multiplication, and $\oplus$ denotes the broadcasting addition.

C. COPYPASTE MODULE

To mitigate the risk of model overfitting, the network initiates image enhancement procedures before engaging in training using the designated dataset. In contemporary practice, image enhancement stands as an effective strategy to augment dataset diversity, and a multitude of enhancement methods are currently available. One prevalent enhancement approach is the application of Random Erasing [29], which encompasses random variation in the length and width of the mask region, along with the pixel substitution values. This technique confers the ability to impart diverse levels of masking to images, thereby engendering robustness across classification, detection, and facial recognition tasks. Another widely employed strategy, known as Cutout [30], functions as a CNN-based regularization technique. It strategically introduces random masking of contiguous image content using square regions, facilitating enhanced model generalization. Confetti [31] serves as an ingenious method for generating synthetic anomalies. It entails the insertion of colored speckles into samples, thereby accentuating the depiction of anomalous local properties. Noteworthy in its specialization, SCADN [17], [18] adopts multi-scale striped masks that span both vertical and horizontal orientations, fostering the aggregation of semantic context to refine detection accuracy. RIAD [21] opts for a distinct approach, selecting to randomly excise sections of the rectangular grid region within the image prior to inputting the sample into the reconstruction network. This technique contributes to the enhancement of anomaly detection by way of restoration-oriented reconstruction.

Anomalies manifest on the surfaces of industrial products due to inherent disparities in the product's properties. We have discerned that overlaying the surface of normal samples with their own minute pixel patches yields an enhancement in network performance. Our aim is to devise an image enhancement strategy by artificially simulating anomalies through this approach. In pursuit of this objective, we introduce a strategy termed CopyPaste, which serves as a mechanism to aid the network in acquiring a deeper understanding of normal semantics.

CopyPaste operates through a process involving the random duplication of small pixel patches, each occupying an area ratio of 1%, within a normal sample. Subsequently, these patches are randomly integrated back into the same normal sample. Notably, the design of CopyPaste hinges on two significant considerations: i) The placement of these small pixel patches onto various areas of the image is rendered equally probable. Given the inherent unpredictability of authentic anomalies, the act of selecting sampling and pasting locations remains arbitrary within this framework. ii) Each individual pixel patch has the capacity to coexist and overlap with others. In this context, we refrain from imposing limitations on the exclusivity of coverage for individual pixels. By embracing these principles, our approach underscores a flexible and unconstrained methodology for enhancing the learning process.

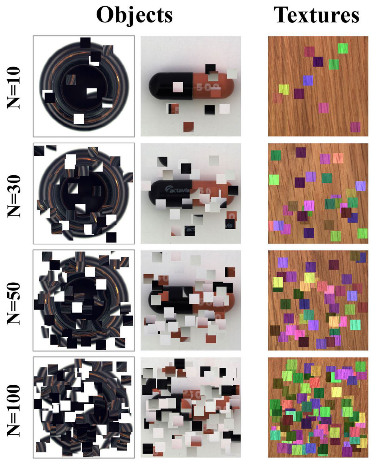


FIGURE 5. The enhancement effect on two types of datasets in MVTec are shown. The number of patches is of size N ∈ {10, 30, 50, 100}. Note that we have added additional color dithering to our treatment of texture-like data.

Figure 5 illustrates the outcomes of applying the CopyPaste strategy to diverse objects. Here, N signifies the count of strategy repetitions. Evidently, as N increases, the patch coverage on the normal image expands correspondingly. Consequently, the network's ability to restore the image becomes increasingly challenging. This deliberate augmentation enhances the unpredictability of image content and augments the informational richness of the image. By training on these enhanced images, the network's capacity for interpreting and recognizing image features is significantly fortified.

Given the uniform distribution of texture class samples, we are focusing our attention on two specific scenarios inherent in texture class images. As demonstrated in Figure 6, we present visualization results exemplifying these scenarios using wood as an illustrative example. The first scenario entails a straightforward application of CopyPaste, without any supplementary operations. In contrast, the second scenario involves the introduction of color dithering to each patch before pasting.

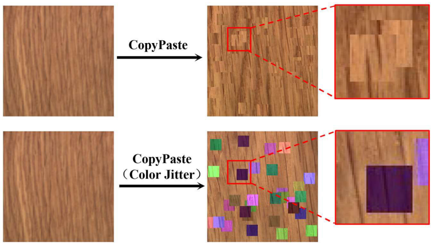


FIGURE 6. CopyPaste enhancement on wood. The first row is direct. The second row is dithered for color.

D. TRAINING OBJECTIVES

To compel the network into a comprehensive grasp of the normal semantics inherent within the training images, we employ a quartet of distinct loss functions: Encoder loss, Content loss, Adversarial loss, and Structural Similarity loss.

1) ENCODER LOSS

Since the outliers for detection are actual data, we train the network model by using the MSE error between the potential vector results Z_1 and Z_2 obtained from two encodings as a supervised signal. Also, to minimise the distance between the bottleneck features of the input image and the reconstructed image, the network uses an encoder loss to force the network to generate reasonable bottleneck features. Its coding loss can be expressed as:

$$\mathcal{L}_{enc} = \frac{1}{n}\sum\nolimits_{i=1}^{n} ||G_{E1}^{i} - G_{E2}^{i}||_2 \tag{4}$$

Where G_{E1}^{i} denotes the encoding result of the ith image after encoder, and G_{E2}^{i} denotes the encoding vector of the i-th image after additional encoder.

2) CONTENT LOSS

In order to improve the network's ability to focus on image content and texture features, the network defines a content loss. The content loss is mainly used to measure the degree of difference and association between the generated image and the real content image. In order to reduce the gap between the two, so that the distribution of the generated image is closer to the distribution of the original image, the loss is expressed as follows:

$$\mathcal{L}_{con} = \frac{1}{n}\sum\nolimits_{i=1}^{n} ||x_i - G(x_i)||_1 \tag{5}$$

Where x_i denotes the ith input image, and $G(x_i)$ denotes the ith generated image.

3) ADVERSARIAL LOSS

The network progressively improves its learning ability through adversarial training. In order to extract essential features from normal samples, the network uses equation (6) as the adversarial loss of the network. This function is based on a modification of the log-likelihood loss and is mainly used to punish very confident miscalculations in the network. Its loss function is represented as follows:

$$\mathcal{L}_{\mathrm{adv}} = -\frac{1}{\mathrm{n}}\sum\nolimits_{\mathrm{i}=1}^{\mathrm{n}} [\log \mathrm{D}(\mathrm{G}(\mathrm{x_i}))] \tag{6}$$

Where $\mathrm{x_i}$ denotes the i-th input image.

4) STRUCTURAL SIMILARITY LOSS

The model built on the above loss function ignores visual inconsistencies, and a Structural Similarity (SSIM) [32] Loss is introduced in order to better measure the visual similarity between two images. SSIM is based on the three aspects of luminance, contrast and structure of the two images to compensate for the visual errors and to comprehensively assess the degree of similarity between image x and y. The quality of the generated image is improved by calculating the structural similarity between the test and the generated image. For given two images x and y, the loss of SSIM can be derived by the following equation:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

$$c_1 = (k_1L)^2,\ c_2=(k_2L)^2 \tag{8}$$

$$\mathcal{L}_{SSIM} = 1 - SSIM(x, y) \tag{9}$$

Where μ_* denotes the mean of *, and σ_*^2 denotes the variance of *, and σ_{xy} denotes the covariance. c_1 and c_2 are constants used to maintain stability, and L denotes the dynamic range of the pixel value. $k_1 = 0.01$ and $k_2 = 0.03$ The advantage of SSIM is that it accelerates the convergence and captures the structural information of the image in a shorter period of time. The value of takes between [0,1], x->1 means the more similar the two are. For better gradient descent and to make the loss function smaller and smaller, the structural similarity loss is defined here as shown in equation (9).

In summary, the total loss of the SCGAN can be expressed as:

$$\mathcal{L}_{total} = \alpha_1\mathcal{L}_{enc} + \alpha_2\mathcal{L}_{con} + \alpha_3\mathcal{L}_{adv} + \alpha_4\mathcal{L}_{SSIM} \tag{10}$$

Among them. $\alpha_1, \alpha_2, \alpha_3$ and α_4 are the hyperparameters of the corresponding loss function.

E. ANOMALY DETECTION

To enhance the effectiveness of the detection task, it is imperative to establish well-defined detection criteria. We adopt the approach outlined in [25] to define the anomaly score (AS) for our model. This anomaly score serves as the yardstick for detecting anomalies by evaluating the disparity between the input sample and its reconstructed counterpart within the discriminator's feature space. The subsequent equation demonstrates that the anomaly score is principally composed of a weighted summation of content loss and encoder loss. Tailored thresholds should be set for distinct datasets, with anomalies being identified when the calculated AS surpasses the designated threshold.

$$AS = \beta\mathcal{L}_{con}+(1-\beta)\mathcal{L}_{enc} \tag{11}$$

Where β denotes the weight parameter in the range 0 - 1, and $\mathcal{L}_{\text{con}}$ and $\mathcal{L}_{\text{enc}}$ denote the content loss and encoder loss of the network model, respectively.

To facilitate meaningful comparisons, we standardized the computed anomaly scores to fit within the [0,1] interval. A greater normalized value indicates an increased probability of the image exhibiting anomalies. The ultimate formulation of the anomaly score is presented below:

$$AS' = \frac{AS - min(AS)}{max(AS) - min(AS)} \tag{12}$$

IV. EXPERIMENTS

A. DATASETS

MVTec [33] serves as an anomaly detection dataset meticulously crafted to emulate authentic industrial inspection settings. It encompasses a total of 5354 high-resolution color images, spanning across 5 distinct texture categories and 10 object categories. Within this dataset, 3629 images are designated for training and validation purposes, while the remaining 1725 images are allocated for testing. The dataset meticulously annotates 73 distinct types of anomalies, encompassing diverse structural variations such as scratches and dents. Notably, MVTec stands out for its faithful emulation of real-world industrial inspection scenarios, accompanied by pixel-level precise annotations of image anomalies. This meticulous labeling serves as a valuable reference point for subsequent anomaly localization investigations. Figure 7 visually showcases several examples of anomaly images within the MVTec.

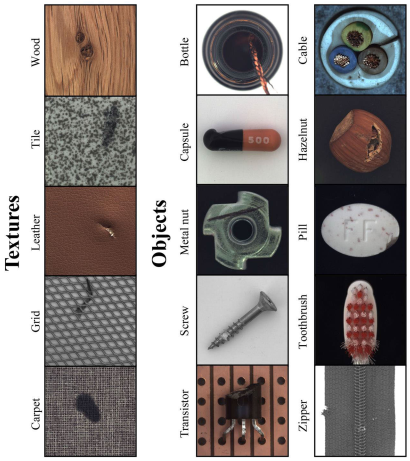


FIGURE 7. Schematic diagram of part of the MVTec dataset.

The Bean Tech Anomaly Detection dataset (BTAD) [26] encompasses a collection of 2830 authentic images depicting various industrial products, capturing both body and surface defects. Comprising RGB images of three distinct industrial products, Product 1 boasts dimensions of 1600 × 1600 pixels, Product 2 measures 600 × 600 pixels, and Product 3 spans 800 × 600 pixels. To harmonize the data, all training images are initially rescaled to 512 pixels before being fed into the model. Each anomaly image is accompanied by a meticulously annotated pixel-level ground truth mask. Illustrated in Figure 8 is a representative sample from the BTAD dataset, showcasing two columns that juxtapose normal and anomalous images for each of the three industrial product types.

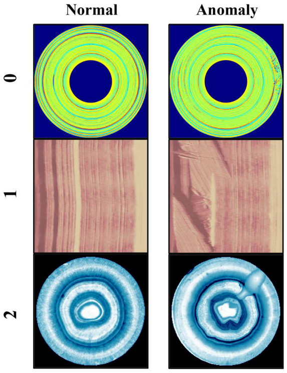


FIGURE 8. Schematic diagram of part of the BTAD dataset.

B. IMPLEMENTATION DETAILS

In this paper, we use the pytorch deep learning framework for anomaly detection and train it using an NVIDIA GTX 3090 GPU with 24 GB of graphics memory. In addition, we use the Adam optimizer [34] to train the network to accelerate network convergence with a learning rate of 0.0002. where $\beta_1 = 0.5$ and $\beta_2 = 0.999$. the number of network training rounds is set to 400 and the batch size is 64. the weights of the loss function are chosen to be $\alpha_1 = 1$. $\alpha_2 = 40$. $\alpha_3 = 1$. $\alpha_4 = 40$.

C. EXPERIMENTAL RESULTS

This paper evaluates network models based on MVTec and BTAD industrial anomaly detection datasets. The evaluation method chooses AUC [35] as a measure, which is a more objective evaluation index of binary classification model, and can comprehensively consider the prediction accuracy and recall of the model. The AUC value tends to be between 0-1, and the larger its value is, the better the classification effect of the network is. Where AUC = 1 indicates a perfect model, and AUC = 0 indicates an invalid model.

Table 1 presents the testing outcomes of AnoGAN [3], GANomaly [25], Skip-GANomaly [36], DAGAN [20], CBiGAN [15], Dual-AttentionGAN [37], and SCGAN on the MVTec dataset. AUC data for AnoGAN, GANomaly, Skip-GANomaly, and DAGAN are sourced from the literature [20]. The data analysis within the table highlights that our proposed

TABLE 1. Surface defect detection performance based on the MVTec open source dataset. For comparison, we report the AUC of seven networks. Data underlined and bolded in the table are the optimal values for each type of test result.

Category		AnoGAN	GANomaly	Skip-GANomaly	DAGAN	CBiGAN	Dual-AttentionGAN	SCGAN
Texture	Carpet	37.7	82.1	79.5	90.3	55.0	91.0	**97.0**
	Grid	87.1	74.3	65.7	86.7	**99.0**	94.0	96.3
	Leather	45.1	80.8	90.8	94.4	83.0	**95.0**	94.7
	Tile	40.1	72.0	85.0	96.1	91.0	80.0	**97.4**
	Wood	56.7	92.0	91.9	97.9	95.0	95.0	**100**
	Average	53.3	80.2	80.2	93.1	84.6	91.0	**97.1**
Object	Bottle	80.0	79.4	93.7	**98.3**	87.0	94.0	**98.3**
	Cable	47.7	71.1	67.4	66.5	81.0	88.0	**98.2**
	Capsule	44.2	72.1	71.8	68.7	56.0	**85.0**	83.2
	Hazelnut	25.9	87.4	90.6	**100**	77.0	95.0	97.5
	Metal nut	28.4	69.4	79.0	81.5	63.0	69.0	**90.1**
	Pill	71.1	67.1	75.8	76.8	81.0	89.0	**89.4**
	Screw	10.0	**100**	**100**	**100**	58.0	**100**	**100**
	Toothbrush	43.9	70.0	68.9	95.0	94.0	**100**	**100**
	Transistor	69.2	80.8	81.4	79.4	77.0	88.0	**91.3**
	Zipper	71.5	74.4	66.3	78.1	53.0	91.0	**92.4**
	Average	49.2	77.2	79.5	84.4	72.7	89.9	**94.0**
Average		50.6	78.2	80.5	87.3	76.7	90.2	**95.1**

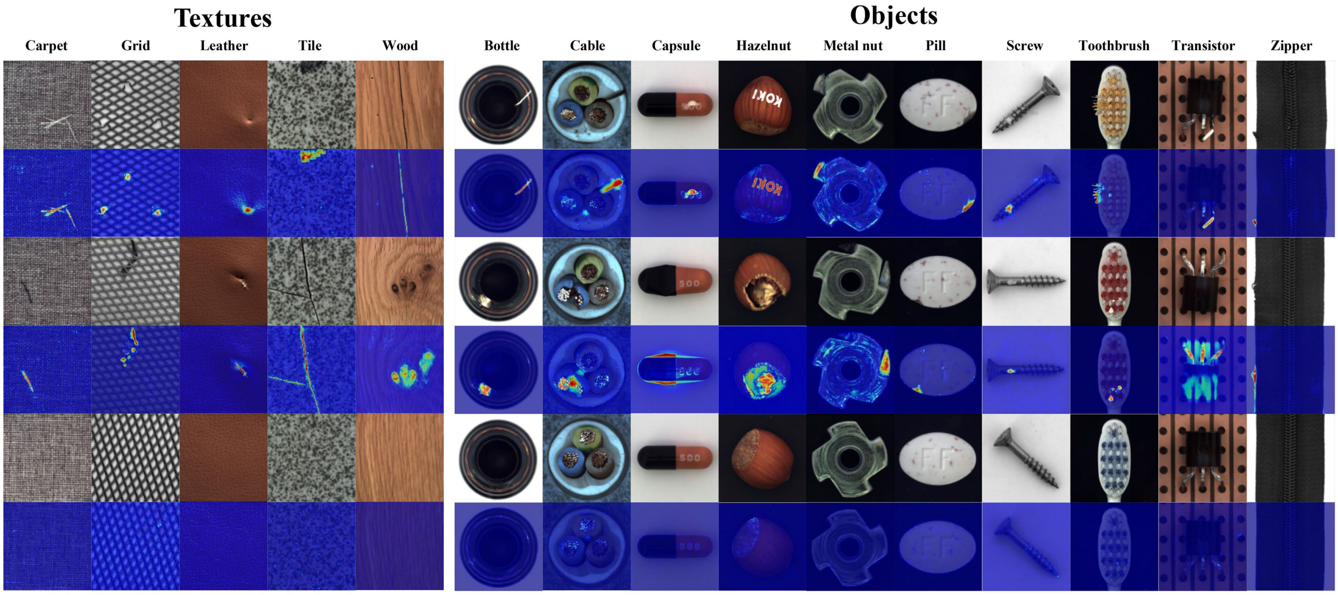


FIGURE 9. Visualization of the anomaly detection results for some MVTec. Rows 1 to 4 mainly show the experimental results for the anomaly images and their thermograms. Row 5 represents the normal image and row 6 shows the abnormal thermogram corresponding to the normal image.

method achieves the highest combined average. Notably, in comparison to Dual-AttentionGAN, our model demonstrates a noteworthy improvement of 4.9 percentage points in the average AUC value. Specifically, SCGAN shows a 6.1% enhancement on the texture class and a 4.1% improvement on the object class. Of significant importance, our model attains a perfect 100% AUC for test results across the wood, screw, and toothbrush.

Figure 9 illustrates select visualization outcomes from the experiments. The experiments were conducted on a dataset categorized into texture and object. Rows 1 through 4 display the abnormal images and their respective localizations. Rows 5 and 6 depict the original image alongside its localized version. From Figure 9, it is evident that SCGAN accurately discerns normal images even without explicit annotation of abnormal areas. Notably, SCGAN not only identifies abnormalities in product images with surface defects but also precisely pinpoints the anomaly locations based on the calculated abnormality probability. These experimental findings underscore the efficacy of SCGAN in detecting surface quality issues in industrial product images. Furthermore, these results offer valuable insights for real-world production quality assessment.

To showcase the robustness of our proposed network, we conducted similar tests using the BTAD dataset. The results of the tests for AE(MSE), AE(MSE+SSIM), VT-ADL, and SCGAN are presented in Table 2. Among these, the experimental data for AE (MSE), AE(MSE+SSIM), and VT-ADL were acquired from the referenced literature [26]. Upon reviewing the table data, it is evident that VT-ADL exhibits a relatively favorable detection performance, achieving an accuracy of 90%. However, our SCGAN model surpasses VT-ADL by further enhancing the detection accuracy by 6.9 percentage points.

TABLE 2. Detection performance based on the BTAD datasets. For comparison, we report the AUC% for the four networks. Data underlined and bolded in the table are the optimal values for each type of test result.

Class	AE (MSE)	AE (MSE+SSIM)	VT-ADL	SCGAN
0	49.0	53.0	99.0	**99.8**
1	92.0	96.0	**94.0**	93.8
2	95.0	89.0	77.0	**97.1**
Mean	78.7	79.3	90.0	**96.9**

V. ABLATION STUDY

To showcase the effectiveness of our proposed method, we conducted ablation experiments across three distinct scenarios. Firstly, we investigate the effect of hyper-parameters on detection performance. Secondly, we empirically demonstrated the influence of the number of CopyPaste patches on detection performance. Thirdly, we delved into the impact of MS-CAM and SSIM within the network. Lastly, we conducted a comprehensive comparative analysis between CopyPaste and various other enhancement strategies.

A. HYPERPARAMETERS

In order to optimize the model's hyperparameters, we conducted an exploration of the influence of various hyper-parameters, as defined in Equation (10), on the model's overall performance. Figure 10 provides a visual representation of the experimental results for these different hyperparameters on the MVTec dataset. It is observed that when $\alpha_1 = 1, \alpha_2 = 40, \alpha_3 = 1$ and $\alpha_4 = 40$, the model attains the highest AUC score. Under these settings, the model effectively captures the semantic characteristics of the samples, thereby enhancing its capability for accurate image anomaly detection.

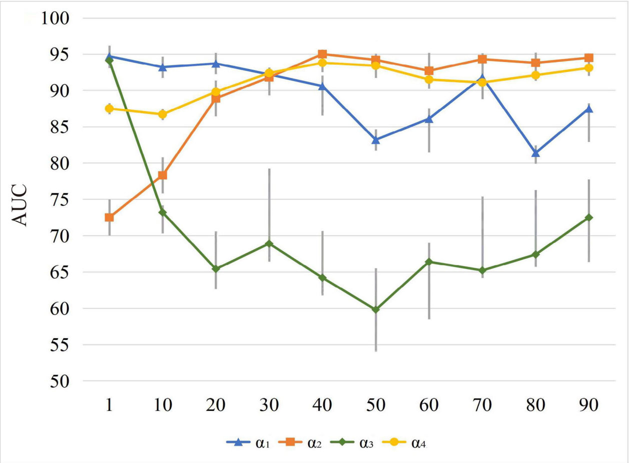


FIGURE 10. Hyper-parameter tuning for the model. The model achieves the most optimum performance when α_1= 1, α_2= 40, α_3= 1, and α_4= 40.

To examine the impact of hyper-parameters on anomaly scores (*AS*) in the experiment, we initiate a discussion regarding the value of β. The anomaly score is defined in Equation (11), with the value of β chosen from the 0-1 interval to represent its magnitude in the experiment. Fig. 11 illustrates the experimental line chart of anomaly scores on MVTec. The findings indicate that the network achieves optimal performance and effectiveness when β is set to 0.8.

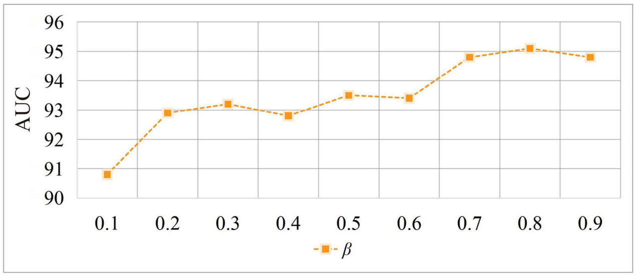


FIGURE 11. Results of the anomaly score experiment. When β = 0.8, the network performance is optimal.

B. NUMBER OF COPYPASTE

While CopyPaste exhibited promising performance in our experiments, the impact of the number of patches on the experimental outcomes remains unclear. To address this, we conducted ablation experiments on the MVTec dataset using CopyPaste with varying numbers of patches. Specifically, we conducted experiments with N values of 10, 30, 50, and 100. The corresponding results are presented in Table 3.

TABLE 3. CopyPaste ablation experiments on the MVTec covering different number of patches. Data underlined and bolded in the table indicate the optimal values for each type of case.

Category		N=10	N=30	N=50	N=100
Texture	Carpet	94.4	**97.0**	96.2	94.4
	Grid	95.6	96.3	**96.8**	94.1
	Leather	92.3	**94.7**	89.9	87.4
	Tile	**97.7**	97.4	95.9	91.1
	Wood	98.8	**100**	99.6	94.7
	Average	95.8	**97.1**	95.7	92.3
Object	Bottle	**99.2**	98.3	93.5	98.6
	Cable	96.4	**98.2**	95.6	93.4
	Capsule	85.1	83.2	85.6	**86.4**
	Hazelnut	96.6	**97.5**	93.2	88.2
	Metal nut	83.5	90.1	**90.4**	86.2
	Pill	92.0	89.4	**92.2**	87.3
	Screw	**100**	**100**	96.5	95.0
	Toothbrush	98.8	**100**	95.0	93.1
	Transistor	**92.4**	91.3	91.1	87.7
	Zipper	89.6	**92.4**	92.2	89.5
	Average	93.4	**94.0**	92.5	90.5
Average		94.2	**95.1**	93.6	91.1

The table data highlights that the model's detection performance suffers when N=100. This is attributed to excessive enhancement, causing the patches to obscure the original semantic features of the image. Essentially, the network loses grasp of half the image's dimensions, hindering comprehensive feature learning. A comparison of the remaining three cases indicates similar detection outcomes for N=10 and N=50, registering at 94.2% and 93.6% respectively. Notably, our model's performance peaks at N=30, achieving an average AUC value of 95.1%. This is because when N=30, the promotion and inhibition effects of CopyPaste on the network reach a balance.

C. MS-CAM & SSIM

This section elucidates the impact of individual structures on the experimental outcomes through ablation experiments. We devised four distinct structures by manipulating the attention module and SSIM loss function within the network. The specific choices for MS-CAM and SSIM are outlined in Table 4.

TABLE 4. Options for four different network architectures.

Class	Struc1	Struc2	Struc3	Struc4
MS-CAM			√	√
SSIM		√		√

As can be seen in Table 5, different network structures achieved different detection results. Struc4 achieved the best average AUC for a single class and the best in the combined evaluation. Struc2 achieved the second best. The experimental results of the four different structures on MVTec illustrate the importance and effectiveness of the combined action of MS-CAM and SSIM.

TABLE 5. AUC values on the MVTec with different sub-modules removed as an indication of the effectiveness of the network improvements. Data underlined and bolded in the table indicate the optimal values for each type of test result.

Category		Struc1	Struc2	Struc3	Struc4
Texture	Carpet	86.5	94.2	94.9	**97.0**
	Grid	**96.4**	95.6	94.7	96.3
	Leather	87.8	89.9	**94.9**	94.7
	Tile	96.5	95.4	91.0	**97.4**
	Wood	96.4	98.4	96.8	**100.0**
	Average	92.7	94.7	94.5	**97.1**
Object	Bottle	95.1	97.0	92.3	**98.3**
	Cable	93.2	92.1	95.8	**98.2**
	Capsule	83.4	**85.5**	87.9	83.2
	Hazelnut	93.6	94.1	89.1	**97.5**
	Metal nut	86.3	88.4	**91.1**	90.1
	Pill	85.2	**91.8**	86.3	89.4
	Screw	94.5	94.6	97.5	**100.0**
	Toothbrush	**100.0**	**100.0**	**100.0**	**100.0**
	Transistor	88.2	92.6	89.9	**91.3**
	Zipper	87.5	87.4	90.4	**92.4**
	Average	90.7	92.4	92.0	**94.0**
Average		91.4	93.1	92.8	**95.1**

Figure 12 depicts the schematic of surface anomaly detection on bottles and grids under four different structures. The first column of the figure displays the input test image, while the second column showcases the ground truth of the anomalies. Subsequent columns from the third to the last exhibit the heat maps of the detection outcomes produced by the four networks. The heat map provides visualization of the abnormal location and size within the test image. Evidently, from Figure 12, the detection effectiveness gradually increases from left to right. The experimental findings indicate that Struc4 yields the most accurate and finely detailed detection results, closely approximating the ground truth.

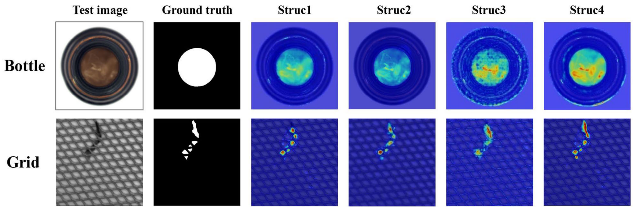


FIGURE 12. Heatmap of some of the detection results. The colors in the graph from blue to red indicate the degree of anomalies in the image from small to large.

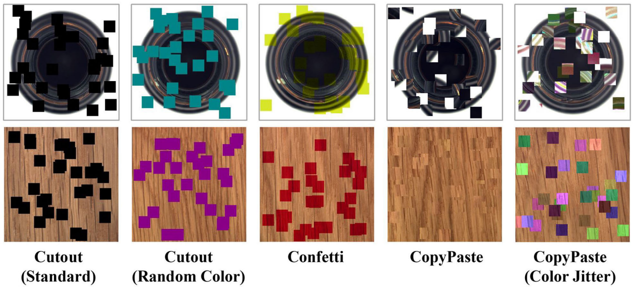


FIGURE 13. Enhancement results for different methods. The first row is lens. The second row is the wood. The square patches in the figure are small pixel areas used for enhancement.

D. ENHANCEMENT METHODS

To compare the various enhancement methods, Figure 13 visualizes five enhancement scenarios for the three methods CopyPaste, Cutout and Confetti: Cutout (Standard) enhances only the black patches; Cutout (Random Color) patches use random color blocks; Confetti inserts colored patches; CopyPaste(Color Jitter) is a color jitter superimposed on a normal object block.

Table 6 shows the experimental results of the above three methods with different enhancement strategies. It can be seen that Cutout's detection is very poor for both black fill and color fill. Confetti synthetic anomaly improves the detection performance. Most importantly, the CopyPaste enhancement strategy resulted in a significant improvement in network performance in this task. It is worth mentioning that CopyPaste

TABLE 6. Ablation experiments with different data enhancements.

Category	Cutout	Cutout (Color)	Confetti	CopyPaste	CopyPaste (Color)
Texture	65.1	66.4	85.7	91.9	**97.1**
Object	76.8	77.9	82.5	**94.0**	92.1
All	73.2	75.3	83.9	92.6	**94.4**

with color dithering attached performs well on texture-like data.

VI. CONCLUSION

Our objective is to extract features solely from normal semantics to facilitate unsupervised anomaly detection. To enable the network to comprehensively capture the normal semantics of industrial product images, we introduce a novel enhancement technique termed CopyPaste. By leveraging the randomness of copy-paste, CopyPaste enhances the model's robustness effectively. Moreover, we incorporate a multi-scale channel attention module into the encoder-decoder-encoder-based generative adversarial network. Experimental results using two real-world datasets underscore the superior anomaly detection performance of our proposed method. In our future work, we plan to integrate CopyPaste as an image augmentation module into a wider array of models to enhance their feature extraction capabilities. Furthermore, we are dedicated to expanding the methodology presented in this paper to encompass anomaly detection in diverse application scenarios.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[2] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student–teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4182–4191.

[3] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 146–157.

[4] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[5] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6479–6488.

[6] R. Wu, S. Li, C. Chen, and A. Hao, "Improving video anomaly detection performance by mining useful data from unseen video frames," *Neurocomputing*, vol. 462, pp. 523–533, Oct. 2021.

[7] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, Mar. 2020.

[8] M. Adiban, S. M. Siniscalchi, and G. Salvi, "A step-by-step training method for multi generator GANs with application to anomaly detection and cybersecurity," *Neurocomputing*, vol. 537, pp. 296–308, Jun. 2023.

[9] X. Tao, W. Ma, Z. Lu, and Z. Hou, "Conductive particle detection for chip on glass using convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[10] H. Di, X. Ke, Z. Peng, and Z. Dongdong, "Surface defect classification of steels with a new semi-supervised learning method," *Opt. Lasers Eng.*, vol. 117, pp. 40–48, Jun. 2019.

[11] X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "GAN-based anomaly detection: A review," *Neurocomputing*, vol. 493, pp. 497–535, Jul. 2022.

[12] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Proc. Artif. Neural Netw. Mach. Learn. (ICANN), 21st Int. Conf. Artif. Neural Netw.*, Espoo, Finland. Berlin, Germany: Springer, Jun. 2011, pp. 44–51.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[15] F. Carrara, G. Amato, L. Brombin, F. Falchi, and C. Gennaro, "Combining GANs and AutoEncoders for efficient anomaly detection," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 3939–3946.

[16] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017, pp. 1–18.

[17] X. Yan, H. Zhang, X. Xu, X. Hu, and P.-A. Heng, "Learning semantic context from normal samples for unsupervised anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 4, pp. 3110–3118.

[18] X. Zhang, S. Li, X. Li, P. Huang, J. Shan, and T. Chen, "DeSTSeg: Segmentation guided denoising student–teacher for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3914–3923.

[19] M. Z. Zaheer, J.-H. Lee, M. Astrid, and S.-I. Lee, "Old is gold: Redefining the adversarially learned one-class classifier training paradigm," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14171–14181.

[20] T.-W. Tang, W.-H. Kuo, J.-H. Lan, C.-F. Ding, H. Hsu, and H.-T. Young, "Anomaly detection neural network with dual auto-encoders GAN and its industrial inspection applications," *Sensors*, vol. 20, no. 12, p. 3336, Jun. 2020.

[21] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107706.

[22] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9659–9669.

[23] J. Pirnay and K. Chai, "Inpainting transformer for anomaly detection," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2022, pp. 394–406.

[24] F. Ye, C. Huang, J. Cao, M. Li, Y. Zhang, and C. Lu, "Attribute restoration framework for anomaly detection," *IEEE Trans. Multimedia*, vol. 24, pp. 116–127, 2022.

[25] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. 14th Asian Conf. Comput. Vis. (ACCV)*, Perth, WA, Australia. Cham, Switzerland: Springer, Dec. 2019, pp. 622–637.

[26] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, "VT-ADL: A vision transformer network for image anomaly detection and localization," in *Proc. IEEE 30th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2021, pp. 1–6.

[27] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2893–2901.

[28] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3559–3568.

[29] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 13001–13008.

[30] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.

[31] P. Liznerski, L. Ruff, R. A. Vandermeulen, B. J. Franks, M. Kloft, and K.-R. Müller, "Explainable deep one-class classification," 2020, *arXiv:2007.01760*.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[33] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[35] C. X. Ling, J. Huang, and H. Zhang, "AUC: A statistically consistent and more discriminating measure than accuracy," in *Proc. IJCAI*, vol. 3, 2003, pp. 519–524.

[36] S. Akçay, A. Atapour-Abarghouei, and T. P. Breckon, "Skip-GANomaly: Skip connected and adversarially trained encoder–decoder anomaly detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

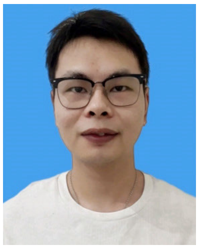
[37] X. Li, Y. Zheng, B. Chen, and E. Zheng, "Dual attention-based industrial surface defect detection with consistency loss," *Sensors*, vol. 22, no. 14, p. 5141, Jul. 2022.

YANG DAI was born in Chengdu, China, in 1997. He is currently pursuing the master's degree with China West Normal University. From 2021 to 2023, a total of four articles were published. His research interests include computer vision and deep learning.

LIN ZHANG was born in Dazhou, China, in 1999. She received the master's degree in electronic information from China West Normal University, in 2023. She is currently a Scientific Researcher with the Sichuan Digital Economy Research Institute. From 2021 to 2023, a total of four articles were published. Her research interests include computer vision and deep learning.

FU-YOU FAN was born in Yibin, China, in 1974. He received the master's degree in computer science from the Chengdu University of Technology, in 2005, and the Ph.D. degree from the University of Electronic Science and Technology of China, China, in 2015. He is currently the Director of the Network and Library Information Center, Yibin University. He has presided over and completed two university-level scientific research projects, participated in two provincial-level scientific research projects, and published more than 40 academic articles. His research interests include artificial intelligence and quantum computing. He is a member of CCF (18135M).

YA-JUAN WU was born in Dazhou, China, in 1974. She received the master's degree in regional economics from the Institute of Regional Economics, Sichuan Normal University, in 2001, and the Ph.D. degree from Sichuan University, China, in 2011. She is currently an Associate Professor with China West Normal University. She has published more than 20 academic articles. Her research interests include image processing and mathematical modeling.



ZE-KUAN ZHAO was born in Nanchong, China, in 1998. He is currently pursuing the master's degree with China West Normal University. From 2021 to 2023, a total of four articles were published. His research interests include computer vision and deep learning.