RESEARCH ARTICLE

FEGAN: A Feature Extraction Based Approach for GAN Anomaly Detection and Localization

FU-YOU FAN[1,2], LIN ZHANG[3], AND YANG DAI[3]
[1]Network and Library Information Center, Yibin University, Yibin 644000, China
[2]Key Laboratory of Intelligent Terminal in Sichuan Province, Yibin 644000, China
[3]Sichuan Digital Economy Research Institute, Yibin 644000, China

Corresponding author: Lin Zhang (linzhang_dy@163.com)

ABSTRACT With the continuous advancement of industry 4.0 and intelligent manufacturing, there is a growing need for automation and intelligence in industrial production processes. Anomaly detection of industrial product surface images is a key technology in achieving this goal and has thus become a significant area of research. However, this endeavor still encounters challenges such as scarcity of abnormal samples, complexities in data labeling, and uncertainties stemming from unknown factors and randomness. To this end, we propose the Feature Extraction-based for Generative Adversarial Network (FEGAN), aimed at detecting and precisely localizing surface anomaly in industrial products. FEGAN focuses on the deep features of an image, and it builds a feature extraction network and an improved generative adversarial network based on VGG19, respectively. We also jointly determines the anomaly score through the deep feature space as well as the Euclidean distance in 2D image space to better identify and locate the anomaly. Furthermore, we introduce a novel Multi-scale Self-Enhancement (M-SE) strategy to bolster the model's generalization capabilities. We conducted training and testing on the MVTec and Bottle-Cap public datasets. A plethora of experimental results indicate that the proposed method outperforms existing methods significantly in terms of anomaly detection. Additionally, through an evaluation of the model's localization accuracy, we demonstrate that FEGAN exhibits certain competitive advantages.

INDEX TERMS Feature extraction, generative adversarial networks, surface anomaly.

I. INTRODUCTION

Ananomaly refers to the presence of abnormal semantic pixels on the surface image of a product, which can occur in any region of any image. Anomaly detection entails identifying and analyzing data points in datasets that deviate from an expected pattern. Currently, anomaly detection finds applications in diverse scenarios, including medical detection [1], video surveillance [2], [3], financial transaction [4], network security [5], industrial materials inspection [6], [7], [8], social network analysis [9], [10], [11], time series analysis [12], [13], [14], and system fault detection [15].

The inspection of surface anomaly on industrial products plays a critical role in ensuring product quality and reliability. It involves a thorough inspection of the product's exterior surface to identify any defects, flaws or abnormalities that may exist. Traditionally, this is mainly done by manual visual inspection, mechanical contact inspection or assisted by simple electronic devices. However, in a wide range of practical applications, the traditional inspection methods suffer from drawbacks such as slow speed, low precision, large errors, failing to meet modern manufacturing industry needs for efficiency, accuracy, and automation. Currently, researchers are exploring emerging technologies that will enable significant advancements in anomaly detection and localization. Industrial product assembly lines are transitioning from manual to machine intelligence, with breakthroughs occurring in
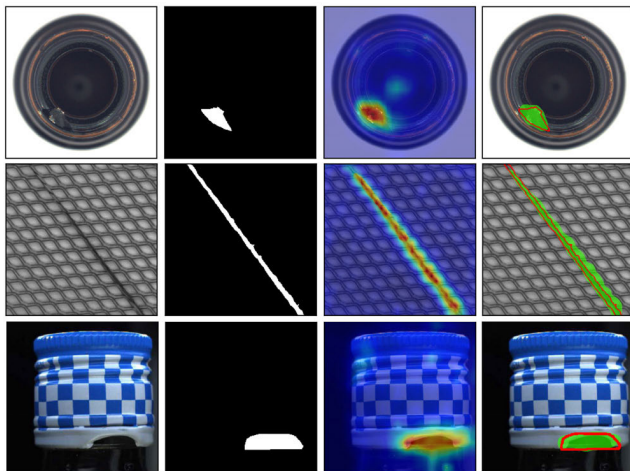
The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

FIGURE 1. Visualization of anomaly detection and localization in MVTec. From left to right are the abnormal image, ground-truth, abnormal superposition thermal map and positioning image respectively.

machine equipment, manufacturing processes, product development, and marketing. As computing power improves and detection algorithms are optimized, the overall quality control of the manufacturing industry is gradually improving.

Image anomaly detection, a crucial technique in computer vision, focuses on identifying anomalous regions in images that deviate from normal patterns. In general, anomaly detection relies on the power of machine learning and deep learning. For most methods, a training set containing a large number of normal images is required for the model to be trained so that it can learn the contextual semantics and feature patterns of the normal samples. When exposed to novel images, the model compares the differences between these images and normal patterns based on its own experience. If the error tolerance is exceeded, the model will identify possible abnormal pixels.

In practical applications, image anomaly detection techniques still face some challenges. Firstly, the definition of anomaly may vary from one application scenario to another, thus requiring the design of appropriate algorithms for different detection tasks. Second, the available anomaly data are often scarcer than normal data, which makes model training more challenging. Third, anomaly patterns are diverse and highly random. How to accurately detect and locate anomaly in the context of strong unknowns is a research challenge.

In this work, we aim to solve the anomaly detection problem on the surface of industrial products, overcome the dependence on anomaly data labeling, and achieve anomaly detection under the condition of only normal training samples. Our research primarily focuses on image feature extraction, aiming to identify unknown anomalies by discerning significant differences between normal and abnormal features. To this end, we propose a GAN anomaly detection and localization method based on feature extraction. The method employs a multi-scale self-enhancement approach to expand the datasets and performs multi-scale feature

matching via dual network construction to detect anomalous pixels in the image. On the one hand, the Feature Extraction Network (FEN) captures the substantial features of the image adhering to the idea of pragmatism. Notably, this contains normal or abnormal features. On the other hand, the Improved Generative Adversarial Network (IGAN) is constructed based on an asymmetric encoder-decoder which understands only the semantic patterns of normal images. The network performs the task of normal reconstruction of abnormal regions under conditions that are insensitive to the distinction between positive and negative class images. We utilize the deep feature extraction function of VGG19 to assist Improved GAN for feature generation and anomaly discrimination, ultimately achieving image-level detection and pixel-level localization of industrial product surface anomaly. Fig. 1 shows the detection and localization effect of our method on MVTec.

In summary, our contributions are as follows:

- We propose a deep learning method based on feature extraction, FEGAN, which is specially used for anomaly detection and location of industrial product surfaces. It mainly consists of feature extraction network and improved generative adversarial network.
- We propose a novel multi-scale self-enhancement method (M-SE) for approximating real anomalies. This further improves the robustness of the network. It copies a multi-scale area of a certain area proportion from the original image and then randomly overlays it on the original image. We control the intensity by setting the area enhancement ratio.
- We jointly determine anomaly from different levels by constructing dual networks. Specifically, in order to make the network focus on deep features and semantic content, we specially designed a novel exception score. It defines the exception criterion from the deep feature space and image space.
- The validity of our proposed model is illustrated by extensive experimental validation and ablation analyses. Excellent performance is shown on both MVTec and Bottle-Cap, two challenging datasets.

II. RELATED WORKS

In the recent literature on anomaly detection, generative adversarial networks [16], [17] and autoencoders [18], [19], [20] have been extensively utilized as deep learning models in various methodologies. Generative adversarial networks strive to mimic real images by continually enhancing their ability to differentiate between genuine and counterfeit images. This adversarial training mechanism contributes to enhancing the model's generalization capacity. Autoencoders learn to efficiently represent images by extracting features and compressing data. Additionally, convolutional neural networks are frequently employed by researchers to construct models based on a reconstruction-based approach.

A. RECONSTRUCTION-BASED APPROACH

Defect-GAN [21] follows the Image-to-Image translation network and relies only on normal samples to execute automated anomaly synthesis tasks. The network performs two-stage step-by-step learning through a defacement process and a recovery process, enabling flexible control over the location and category of generated anomaly. Experiments confirm Defect-GAN's capability to synthesize diverse anomaly with a considerable degree of fidelity, thereby enhancing the network's performance in anomaly detection on concrete bridge surfaces. DRAEM [22] restores anomalous regions featuring semantically meaningful anomaly through a reconstruction sub-network, and then connects the outputs of the reconstructive sub-network with the input images and feeds them into a discriminative sub-network with skip-connections. Niu et al. [23] proposed a controllable anomaly image generation method concerning region and intensity. They constructed anomaly direction vectors in the potential space based on feature continuity on both sides of the anomaly boundary to control anomaly intensity. Moreover, the model is guided to focus on the anomalous region through construction of anomalous attention loss. Gao et al. [24] employ GAN to generate reconstruction of low-quality images with masks, subsequently utilizing the VGG16 network for anomaly identification. Yan et al. [25] propose a semantic context-based anomaly detection network called SCADN. Notably, this network randomly overlays the input image with strip masks of different scales, widths, and orientations, followed by network reconstruction. OGNet [26] constructs a two-stage training task and uses the discriminator to measure the reconstruction quality of the image. It is worth mentioning that OGNet treats as an enhanced anomaly image. RIAD [27] transforms anomaly detection into an image recovery problem by randomly removing sub-regions in the image and then performing recovery reconstruction based on the neighborhood. Sabokrou et al. [28] proposed R and D networks to enhance the framework's performance through adversarial training while augmenting original data and suppressing anomalous data. Skip-GANomaly [29] introduces skip-connections to preserve both global and local features. However, the computational complexity of the model is high and relies heavily on the quantity and quality of training data.

B. EMBEDDING-BASED APPROACH

This approach assumes that a sample as abnormal if its features significantly deviate from the normal feature distribution. PaDim [30] relies on a pre-trained CNN feature extractor to obtain a probabilistic representation of a normal image through a multivariate Gaussian distribution. Yang et al. [31] introduce normal memory samples and anomalous samples to assist in MemSeg learning, utilizing spatial attention and multiscale feature fusion modules in the model. MemSeg demonstrates high accuracy in anomaly detection tasks and exhibits a robust inference advantage with a processing speed of 0.0319 seconds per image. SPADE [32]

performed anomaly detection by aligning the test image with the retrieved normal image and segmentation. The method relies on the K nearest neighbors of the pixel-level feature pyramid. Patch SVDD [33] extends the traditional support vector data description method to the patch level for anomaly detection and segmentation. Bergmann et al. [34] propose a ST detection framework by comparing the differences in sample embedding between the teacher network and the student network, as well as the anomalous sample embedding differences to identify anomalous samples in the data.

III. PROPOSED METHOD

In this section, we will begin by providing an overview of the overall model and its staged process. Subsequently, we will delve into the feature extraction network and the improved generative adversarial network separately. Finally, we will present additional details of this method.

A. FEGAN

Figure 2 demonstrates our proposed network framework and algorithm flow. Overall, the model mainly consists of a Feature Extraction Network (FEN) and an Improved Generative Adversarial Network (IGAN). The FEN adopts the structure of VGG19 [35] following pre-training on ImageNet [36]. It serves the purpose of extracting high-level feature representations, including overall shape and contextual semantics, from an image. On the other hand, the IGAN consists of a generator and a discriminator, where the generator consists of an encoder and a decoder. Here the encoder is also used as a feature encoding network.

The training phase, depicted by the green solid line in Fig. 2. The network receives only normal image samples as input. Normal images without data enhancement are fed into the pre-trained VGG19 for extracting depth features. Subsequently, the enhanced normal image is processed through M-SE image enhancement before being fed into the generator for sample reconstruction. In the next step, the reconstructed images, along with the original normal images, are then inputted into the discriminator for adversarial training.

Theoretically, the FEN is able to extract the essential features for all types of images, both normal and abnormal. However, the feature encoding network, having been trained only on normal image features, theoretically reconstructs abnormal images as normal ones after self-enhancement. Consequently, the feature loss and content loss between the two types of images should be substantial.

The test phase, represented by the red solid line in Fig. 2, operates differently depending on whether a normal or abnormal image is inputted. For normal images, the depth feature residuals of FEN and the feature encoder are minimal, resulting in a reconstructed image similar to the original. Therefore, the output anomaly score falls below the anomaly threshold. Conversely, for abnormal image inputs, FEN extracts abnormal features, which differ significantly from the normal features extracted by the encoder. This leads to an output anomaly score surpassing the anomaly threshold. Finally,
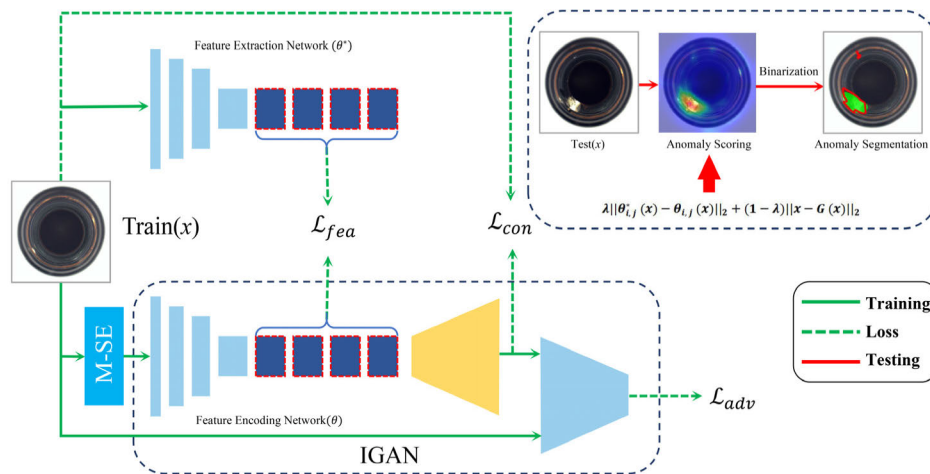
FIGURE 2. Network structure of FEGAN. The training flow is shown as green solid line. The testing flow is shown as the red solid line. The green dashed line shows the composition of the loss function.

in the dotted box section, we show the anomaly score. Based on this, the anomaly detection heat map is output and anomaly segmentation is achieved by binarization.

B. FEATURE EXTRACTION NETWORK

In order to utilize image depth features for anomaly detection, inspired by Yang et al. [37], we construct a feature extraction network based on VGG19. This network can effectively capture high-level feature representations, encompassing both normal and abnormal features, for any input image.

Overall, VGG19 uses a five-segment convolution structure. And three fully connected segments and softmax layers are attached to it. In detail, the VGG19 contains a total of 19 layers, which are sixteen convolutional layers and three fully connected layers. In terms of parameter settings, the VGG19 network uses the 3×3 convolution kernel as well as 2×2 the max-pooling layer. This reduces the number of network parameters while preventing model overfitting. Finally, three FC layers are used to output the class probabilities of the images.

In the convolution process, the first convolution block includes two conv layers and a max-pooling layer, primarily tasked with extracting low-level features like color and texture from the image. This lays the groundwork for subsequent image tasks such as feature extraction. The second convolution block mirrors the structure of the former, aiming to continue extracting basic image features while enhancing the network's capability to extract higher-level features. The subsequent two convolution blocks further build upon the first two, extracting intermediate-level features such as texture combinations and local shapes. Finally, the last convolution block focuses on extracting high-level key features such as overall shape and contextual semantics. As the number of convolution layers increases, the depth and complexity of the network also increases, gradually achieving abstract feature extraction of the image. This hierarchical extraction substantially improves the network performance.

Since VGG19 was pretrained on the large ImageNet datasets, the network demonstrates good generality for normal/abnormal feature extraction. It's worth noting that FEGAN only utilizes features extracted from the fifth convolution block. The FEN offers a robust representation of deep features for network training, which is indicative of anomaly detection and localization.

C. IMPROVED GENERATIVE ADVERSARIAL NETWORK

Improved Generative Adversarial Network (IGAN) significantly enhance the extraction of key image features. The feature encoding network within IGAN plays a pivotal role in determining the quality of subsequent synthetic images, forming the foundation for anomaly detection tasks.

The generator (G) is introduced first. It comprises a feature encoding network and a decoder, aiming to extract deep convolution features before reconstructing them. The structure of the feature encoding network closely resembles that of VGG19, but in our model, we replace max-pooling with average-pooling. This change is inspired by literature [37] and ensures that the network retains more comprehensive information, crucial for anomaly detection and segmentation. This is because max-pooling only retains the maximum value in the convolution kernel region and ignores other parts of the information, which may cause the network to lose certain anomalous region information, thus affecting the segmentation performance of the network. Additionally, we discard the fully connected layer in VGG19. FEGAN capitalizes on the deep feature extraction prowess of VGG19, enabling the capture of complex and abstract surface image details. After encoding features, the decoder reconstructs feature vectors into the original image space, albeit in a lower dimensionality.

The encoding-decoding process of the generator adopts an asymmetric structure consisting of five convolutional blocks and one deconvolution block, respectively. These blocks achieve deep feature extraction through multiple iterations of convolution. Specifically, the attributes of the input image are 256×256 of a three-channel colour image. The size of

the convolution kernel is 3×3. The size of the transposed convolution kernel is 4×4. The image is reconstructed back to RGB image by transpose convolution.

During the training, the generator learns to represent only normal industrial product surface images, aiming to restore them to their original state after passing through the feature encoding network. In the testing phase, due to its training on normal images exclusively, the generator tends to reconstruct images with anomaly as normal ones, resulting in a noticeable disparity.

The role of the generator focuses on extracting the normal features of the image. Firstly, it lays the foundation for the reconstruction of the decoder and facilitates the subsequent calculation of content loss. Secondly, it is used to calculate the feature loss in conjunction with the feature extractor to facilitate model training.

The discriminator (D), another vital GAN component, opposes the generator by distinguishing between real dataset images and fake generator images. It guides the generator in adversarial learning, improving image reconstruction quality while enhancing its own discriminative ability.

D. MULTI-SCALE SELF-ENHANCEMENT

In this work, we only have normal samples available for reference. At the same time, there is a big difficulty in the collection of real anomaly. Recognizing this limitation, we aim to enhance the robustness of FEGAN by introducing enhancement mechanism. We found that most of the current enhancement methods [25], [27], [38] cover part of the region with masks or images such as squares, rectangles, and stripes. Their enhancement of the image follows a logic of uniformly using squares, rectangles, or stripes. However, the appearance of anomaly is often irregular and random. Therefore, we propose a novel method called Multi-scale Self-Enhancement (M-SE).

M-SE takes into account the randomness and irregularity of anomaly appearing on the surface of a product. The core concept involves copying a certain area proportion of the multi-scale region from the original image, and then paste it to an arbitrary position of the original image. There are several features of M-SE that are worth paying attention to: firstly, we stipulate that the copying proportion is controlled by the area. For example, specifying that the copied area is 30% of the original image means that it is randomly sampled from 30% of the area and randomly pasted. This is due to the uncertainty of the anomaly. Here, the size of the sampling area is equal to the sum of the enhanced areas. Secondly, we adopt multi-scale anomaly enhancement, allowing for variability in the length and width of the enhanced region. This has the advantage of being closer to the real anomaly. Thirdly, we constrain the area of a single patch to occupy no more than 1% of the original image, encouraging FEGAN to focus on detecting tiny abnormalities. The visual effect of M-SE is depicted in Fig. 3.

The M-SE strategy utilizes the normal image itself as the enhancement object and randomly alters its color attributes.
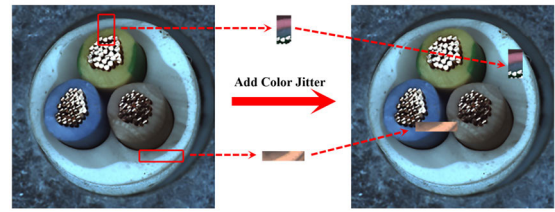

FIGURE 3. Schematic diagram of Multi-scale Self-Enhancement. The method takes patches of any aspect ratio from its own image and inserts them into the original image after adding colour dithering.

This approach offers several advantages. On the one hand, it expands the size of the training dataset, meeting the model's requirement for a large and diverse dataset. On the other hand, it diminishes FEGAN' s reliance on specific attributes within the training dataset.

E. TRAINING OBJECTIVES

The loss function is used to measure the gap between the real data and the predicted data, which is theorized to have a large loss value between normal and abnormal images. To motivate the network training, we designed feature loss, content loss and adversarial loss. The details of each will be developed below.

Feature Loss: In order to force the feature extraction network and the encoding network to align high-level semantic information in the hidden space, we introduce feature loss to focus on sample pixel similarity and feature space consistency. $\mathcal{L}_{fea}$ describes the L_2 distance between the fifth feature module $\theta^*(x)$ of the VGG19 network and the corresponding position of the fifth feature module $\theta(x)$ of the encoding network in G. We improve the quality and diversity of the generated images during training by minimizing, making the reconstructed images closer to the image patterns of the real datasets. In the testing, since the abnormal features behave differently from the normal ones, the $\mathcal{L}_{fea}$ should have larger values. The feature loss is defined as shown in the following equation:

$$\mathcal{L}_{fea} = \mathop{\mathbb{E}}_{x \sim p_x} ||\theta^*(x) - \theta(\tilde{x})||_2 \qquad (1)$$

$$\tilde{x} = (1 - \mathrm{M}) \odot x + \mathrm{M} \odot x_{\mathrm{M-SE}} \qquad (2)$$

where x denotes the original input image. $\tilde{x}$ denotes the image after M-SE enhancement strategy. $x_{\mathrm{M-SE}}$ is a patch image in the x input image region of the patch image that performs M-SE. M represents the mask of M-SE. $\odot$ represents element-by-element multiplication.

Content Loss: The above feature loss focuses on extracting the deeper features of the image. To improve FEGAN' s focus on the overall content features of the image and to maintain the stylistic consistency between the generated image and the original image, we use the content loss to measure the difference between the two in terms of image content, including information such as low-level pixels and high-level semantics. By minimizing $\mathcal{L}_{con}$, our network is able to produce more interpretable visual representations that are

consistent in terms of colour, texture, shape, and structure. The reconstructed images tend to be more natural and realistic. During the testing, since the anomalous images possess different content representation modes, the $\mathcal{L}_{con}$ should have a larger value. The content loss formula can be expressed as follows:

$$\mathcal{L}_{con} = \mathop{\mathbb{E}}_{x \sim p_x} ||x - G(\tilde{x})||_1 \qquad (3)$$

Adversarial Loss: Considering content consistency while avoiding problems such as pattern collapse. Adversarial loss is a key element of GAN network for adversarial game training, and is also a core component of the loss function, which is directly related to the dynamic learning process of the network. During the training, the $\mathcal{L}_{adv}$ pushes generator to explore more potential spaces. The adversarial loss is calculated as follows:

$$\mathcal{L}_{adv} = \mathop{\mathbb{E}}_{x \sim p_x} [logD(x)] + \mathop{\mathbb{E}}_{x \sim p_x} [log(1 - D(G(\tilde{x})))] \qquad (4)$$

In summary, the overall loss of our network is shown in Eq. (5), where λ_{fea}, λ_{con} and λ_{adv} are the weighting parameters.

$$\mathcal{L}_{total} = \lambda_{fea}\mathcal{L}_{fea} + \lambda_{con}\mathcal{L}_{con} + \lambda_{adv}\mathcal{L}_{adv} \qquad (5)$$

F. ANOMALY SCORE
In order to better evaluate the image quality and assist the network to perform the anomaly detection task, we deliberately define the anomaly score by combining the feature space and image space. From the feature point of view, theoretically the depth features of abnormal image and normal image should have a big gap because of their essential differences. Therefore, the distance between the fifth feature module of the extraction network and the coding network is calculated as $F_{i,j}(x)$. From the perspective of image, there should be a large gap between the reconstructed image and the input abnormal image due to different distribution modes, so the L_2 distance between the two is calculated to obtain $R_{i,j}(x)$.

$$F_{i,j}(x) = ||\theta_{i,j}^*(x) - \theta_{i,j}(x)||_2 \qquad (6)$$
$$R_{i,j}(x) = ||x - G(\tilde{x})||_2 \qquad (7)$$

where x is the input image. $\tilde{x}$ is the enhanced image. (i, j) denotes the feature location. When calculating the anomaly score, $F_{i,j}(x)$ needs to be scaled to the same resolution as the reconstructed image.

In addition, we define pixel-level anomaly scores for accurately performing the anomaly localization task $S_{i,j}(x)$. Eq. (8) integrates the feature space with the image space in a substantial gap. We use the weighted sum of $F_{i,j}(x)$ and $R_{i,j}(x)$ as the basis for anomaly determination. Where λ is the weighting parameter.

$$S_{i,j}(x) = \lambda F_{i,j}(x) + (1 - \lambda)R_{i,j}(x) \qquad (8)$$

In order to standardize the data distribution and improve the model stability, we specially normalize the defined exception scores $S_{i,j}(x)$. FEGAN uniformly converts the data into

the normalized interval of [0, 1], eliminating the scale difference between the data for easy model processing and understanding. The normalization is calculated as follows:

$$S_{i,j}(x)' = \frac{S_{i,j}(x) - min(S_{i,j}(x))}{max(S_{i,j}(x)) - min(S_{i,j}(x))} \qquad (9)$$

Of these, the $max(\cdot)$ and $min(\cdot)$ denote taking the maximum and minimum values, respectively.

For image-level anomaly detection, we take the maximum value of the pixel-level anomaly fraction $S_{i,j}(x)$ as the criterion of anomaly determination. The formula is shown below:

$$S(x) = \max_{i,j} S_{i,j}(x) \qquad (10)$$

IV. EXPERIMENTS
In this section, we provide detailed explanations regarding the datasets utilized in the algorithm, the evaluation metrics employed, and the experimental setup. Furthermore, we present the final experimental results both quantitatively and qualitatively, accompanied by a focused data analysis. Finally, we demonstrate the effectiveness of the algorithm by comparing it with similar methods for the task at hand.

A. DATASETS
To assess the robustness of the proposed method, we choose two challenging datasets, which are MVTec [20] and Bottle-Cap [38].
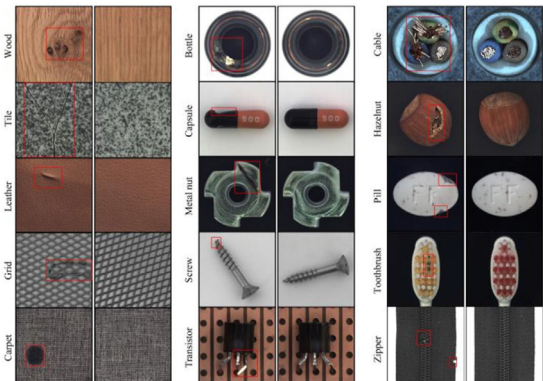

FIGURE 4. Schematic diagram of a portion of the MVTec dataset. The figure shows normal and abnormal images for the 15 classes and the anomalies are labeled with red boxes.

MVTec is a comprehensive anomaly detection dataset designed to simulate real industrial inspection scenarios. It comprises a total of 5354 high-resolution color images. Notably, MVTec covers a wide range of anomaly types, including structural anomaly, morphological anomaly, texture anomaly, and color anomaly, among others. An additional advantage of MVTec is its provision of pixel-level labeling of anomaly, which greatly facilitates subsequent experiments. Examples of images from the MVTec are depicted in Fig. 4.

The Bottle-Cap is an industrial vision inspection data set specifically designed for classification and identification,

TABLE 1. Comparison of image-level anomaly detection performance (AUC-ROC) based on the MVTec AD dataset. The reported results for the first three methods are from the literature [34], and the results for ST-m and DFR are from the literature [30]. Data underlined and bolded in the table are the optimal values for each type of test result.

	Category	GeoTrans	GANomaly	ITAE	DAGAN	ST-m	DFR	SCGAN	FEGAN
	Bottle	0.744	0.892	0.941	0.940	0.994	**0.998**	0.983	0.992
	Cable	0.783	0.760	0.832	0.880	0.790	0.813	0.982	**0.987**
	Capsule	0.359	0.732	0.681	0.850	0.934	**0.971**	0.832	0.955
	Hazelnut	0.670	0.785	0.855	0.950	0.985	**1.000**	0.975	0.989
Object	Metal nut	0.813	0.700	0.667	0.690	0.922	0.939	0.901	**0.981**
	Pill	0.630	0.743	0.786	0.890	0.914	0.930	0.894	**0.962**
	Screw	0.500	0.745	**1.000**	**1.000**	0.828	0.949	**1.000**	0.963
	Toothbrush	0.972	0.653	**1.000**	**1.000**	0.997	0.997	**1.000**	0.984
	Transistor	0.869	0.792	0.843	0.880	0.759	0.817	0.913	**0.945**
	Zipper	0.820	0.745	0.876	0.910	0.929	0.907	0.924	**0.988**
	Carpet	0.437	0.699	0.706	0.910	0.931	**0.976**	0.970	0.961
	Grid	0.619	0.708	0.883	0.940	0.988	0.972	0.963	**0.992**
Texture	Leather	0.841	0.842	0.862	0.950	0.963	0.993	0.947	**0.995**
	Tile	0.417	0.794	0.735	0.800	**1.000**	0.926	0.974	0.987
	Wood	0.611	0.834	0.923	0.950	0.988	0.986	**1.000**	0.992
	Average	0.672	0.762	0.839	0.902	0.928	0.945	0.951	**0.978**

FIGURE 5. Example of a sample of Bottle-Cap. The figure contains anomalies such as deformation, dents, and crushing, and the anomalies are labeled with red boxes.

and each part of it contains carefully labeled images. The Bottle-Cap dataset comprises over 1100 high-resolution color images of bottle caps obtained from real production lines. Unlike MVTec, Bottle-Cap is a specialized dataset focused solely on a single class of images. However, despite its narrower scope, Bottle-Cap encompasses seven distinct real anomalies. A partial example of images from the Bottle-Cap is illustrated in Fig. 5.

B. IMPLEMENTATION DETAILS

In this work, we utilize the PyTorch deep learning framework to construct our model. The hardware environment comprises an Intel i7-12700 and NVIDIA 3090 24GB GPU, while the system operates on Windows 10. The images are scaled to a resolution of 256×256 and the color mode is RGB. The algorithm is trained on normal samples only, using the Adam optimizer with an initial learning rate of 2×10^{-4} and the weight decay of 1×10^{-5}. The weight parameters of loss function $\mathcal{L}_{total}$ are chosen as $\lambda_{fea} = 30$, $\lambda_{con} = 30$, $\lambda_{adv} = 1$.

$\lambda = 0.6$ for anomaly score $S_{i,j}(x)$. Training epoch is set to 200 and batch size is 16.

To effectively assess the performance of our method against similar approaches, we employ the AUC-ROC evaluation metric [39]. AUC-ROC is commonly used to evaluate the performance of both image-level and pixel-level anomaly detection tasks, providing insights into the algorithm's strengths and weaknesses.

Furthermore, we employ the Intersection over Union (IoU) as a secondary evaluation metric to measure the accuracy of the algorithm's localization. IoU quantifies the degree of overlap between the predicted anomalous areas and the true anomalous areas. Higher IoU values indicate closer alignment between the prediction and ground truth. Specifically, an IoU of 1 signifies complete overlap, while an IoU of 0 indicates no intersection between the predicted and ground truth anomalous regions.

C. EXPERIMENTAL RESULTS

To quantitatively analyze the detection and segmentation performance of the FEGAN on the MVTec and Bottle-Cap, we compare the proposed method with seven alternative methods, namely GeoTrans [40], GANomaly [41], ITAE [42], DAGAN [43], ST-m [34], DFR [44], and SCGAN [38]. The anomaly detection results are presented in Table 1.

Image-level AUC-ROC is measured per image. As indicated in Table 1, our proposed algorithm achieves the highest overall average AUC-ROC compared to the aforementioned methods, with a margin of 2.7 percentage points higher than SCGAN. GeoTrans and GANomaly exhibit inferior performance, likely due to their limitations in extracting high-level features from complex images. Specifically, FEGAN demonstrates the best performance for image-level anomaly detection in 7 out of 15 classes. Although it falls short of achieving the optimal performance in classes such as

TABLE 2. Comparison of pixel-level anomaly detection performance (AUC-ROC) based on MVTec AD dataset. The data of AE-ssim and AnoGAN are from literature [12]. The data results of the remaining methods are from the literature [30]. Data underlined and bolded in the table are the optimal values for each type of test result.

	Category	AE-ssim	AnoGAN	VAE-grad	SPADE	ST-m	DFR	FEGAN
Object	Bottle	0.930	0.860	0.922	**0.984**	0.973	0.962	0.982
	Cable	0.820	0.780	0.910	0.972	0.956	0.922	**0.985**
	Capsule	0.940	0.840	0.917	0.990	0.979	0.986	**0.992**
	Hazelnut	0.970	0.870	0.976	**0.991**	0.985	0.986	0.985
	Metal nut	0.890	0.760	0.907	0.981	**0.989**	0.937	0.978
	Pill	0.910	0.870	0.930	0.965	0.979	0.970	**0.982**
	Screw	0.960	0.800	0.945	0.989	0.987	**0.992**	0.991
	Toothbrush	0.820	0.900	0.985	0.979	**0.988**	0.987	**0.988**
	Transistor	0.900	0.800	0.919	0.941	0.782	0.796	**0.968**
	Zipper	0.880	0.780	0.869	0.965	**0.983**	0.973	0.974
Texture	Carpet	0.870	0.540	0.735	0.975	0.951	0.970	**0.989**
	Grid	0.940	0.580	0.961	0.937	0.985	0.980	**0.992**
	Leather	0.780	0.640	0.925	0.976	**0.994**	0.993	0.993
	Tile	0.590	0.500	0.654	0.874	**0.967**	0.913	0.956
	Wood	0.730	0.620	0.838	0.885	0.955	0.954	**0.968**
	Average	0.862	0.743	0.893	0.965	0.964	0.955	**0.982**

TABLE 3. Comparison of anomaly segmentation performance (IoU) based on MVTec AD dataset. The reported results of the comparison methods are taken from the literature [30]. Data underlined and bolded in the table are the optimal values for each type of test result.

	Category	AE-ssim	PatchSVDD	ST-m	DFR	DFC	FEGAN
Object	Bottle	0.293	0.439	0.288	0.253	0.454	**0.579**
	Cable	11.5	30.1	0.232	0.257	0.280	**0.443**
	Capsule	0.094	0.106	0.383	**0.425**	0.328	0.382
	Hazelnut	0.330	0.291	0.399	0.450	0.301	**0.448**
	Metal nut	0.132	0.351	0.388	0.319	**0.449**	0.445
	Pill	0.120	0.161	0.321	**0.363**	0.362	0.362
	Screw	0.136	0.051	0.215	**0.272**	0.170	0.268
	Toothbrush	0.225	0.197	0.239	0.289	0.247	**0.291**
	Transistor	0.166	0.293	0.162	0.186	0.203	**0.368**
	Zipper	0.099	0.293	0.231	0.310	0.310	**0.435**
Texture	Carpet	0.043	0.335	0.238	0.317	0.390	**0.410**
	Grid	0.284	0.173	0.356	0.406	0.410	**0.426**
	Leather	0.023	0.247	0.444	**0.496**	0.474	0.472
	Tile	0.077	0.167	0.288	0.197	0.232	**0.368**
	Wood	0.094	0.184	.247	0.197	0.233	**0.301**
	Average	0.149	0.239	0.295	0.316	0.323	**0.399**

Bottle, Hazelnut, Toothbrush, Tile, and Wood, its AUC-ROC value exceeds 0.98, indicating relatively robust model performance. Conversely, performance in classes such as Capsule, Screw, and Carpet shows room for improvement, reflecting the model's lesser sensitivity to subtle anomaly under image-level detection conditions. Notably, ITAE, DAGAN, and SCGAN achieve 100% AUC-ROC in Screw and Toothbrush classes, suggesting their proficiency in detecting more detailed object class images, an area where our method requires enhancement.

Pixel-level AUC is determined for each individual pixel. To comprehensively evaluate the model's classification ability in local regions, FEGAN is compared with AE-ssim [19], AnoGAN [17], VAE-grad [45], SPADE [32], ST-m, and DFR for pixel-level performance. The experimental results from Table 2 reveal that FEGAN attains an average AUC of 0.982 and excels in performance across five object classes and three texture classes, outperforming the state-of-the-art method

by 1.7 percentage points. AnoGAN suffers from a significant drawback, as it requires constant optimization of latent variables, resulting in the poorest performance on MVTec. SPADE achieves the second-best detection performance with 0.965. Although ST-m ranks third, it delivers the best performance across 5 classes.

To assess the accuracy of FEGAN in anomaly localization on industrial product surface images, we compared the model's anomaly segmentation performance with five alternative methods: AE-ssim, PatchSVDD [33], ST-m, DFR, and DFC [37]. Table 3 provides detailed IoU experimental results for each method on the MVTec.

As shown in the table, FEGAN achieves the highest overall average IoU value of 0.399, surpassing DFC by 0.076. This indicates that the anomalous regions predicted by the model exhibit the greatest overlap with the real regions, resulting in more accurate anomaly localization. Across the 15 classes of the MVTec, FEGAN achieves optimal performance in

TABLE 4. Performance comparison of anomaly detection and segmentation based on Bottle-Cap dataset. The data for the first five methods are from the literature [30]. The data underlined and bolded in the table are the optimal values for each type of test result.

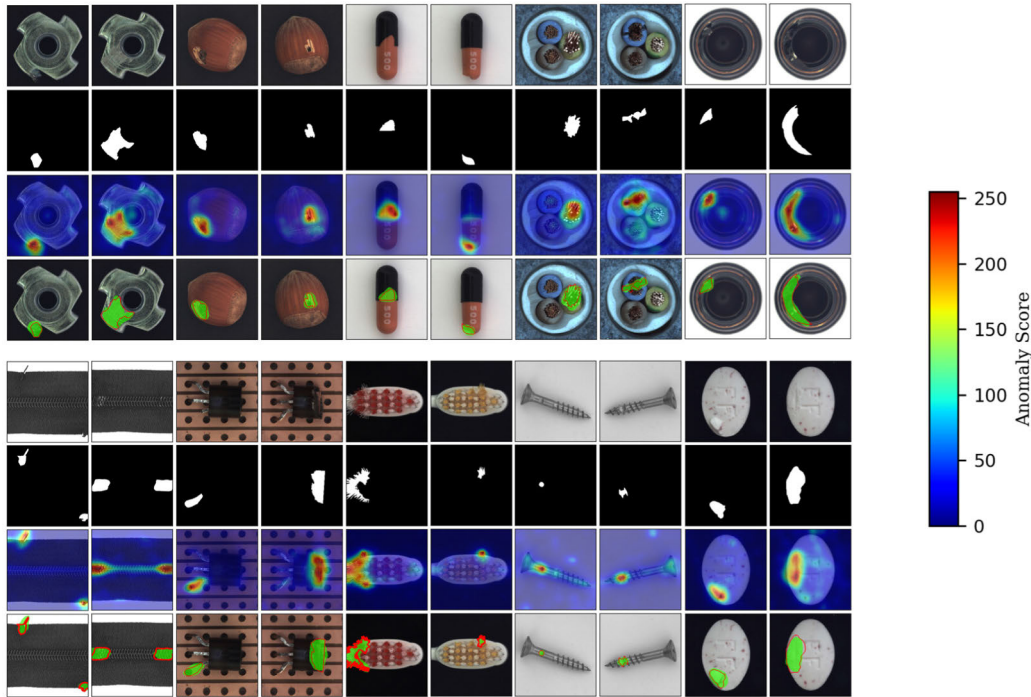| Metircs	AE-ssim	PatchSVDD	ST-m	DFR	DFC	FEGAN
AUC-ROC (image-level)	-	0.649	0.753	0.719	0.941	**0.956**
AUC-ROC (pixel-level)	0.873	0.895	0.946	0.920	0.971	**0.979**
IOU	0.135	0.104	0.191	0.182	**0.289**	0.231

FIGURE 6. Anomaly localization results of FEGAN on object classes of MVTec. It is divided into two parts, top and bottom, and each part has four rows, which are the original anomaly image, the real binarized anomaly image, the anomaly heat map, and the anomaly segmentation map.

10 classes. It can accurately identify and locate abnormal areas. Compared to the other five methods, FEGAN demonstrates superior adaptability in anomaly localization.

In order to evaluate the generalization ability of the proposed model, we compare image-level AUC-ROC, pixel-level AUC-ROC, and IoU performance on the Bottle-Cap respectively. As shown in Table 4, the model demonstrates superior anomaly classification performance on Bottle-Cap, achieving image-level and pixel-level AUC-ROC scores of 0.956 and 0.979, respectively, which are the highest among the six methods evaluated. However, due to the dataset's overall low lighting conditions and the model's decreased performance under such conditions, as well as its sensitivity to large changes in anomaly sizes or complex anomaly shapes, FEGAN ranks second to DFC in anomaly localization. FEGAN achieves an IoU value of 0.231, which is 0.058 lower compared to DFC.

Overall, FEGAN demonstrates strong performance in anomaly detection classification on both datasets, showcasing its ability to classify anomaly at both image and pixel levels by evaluating AUC-ROC at different

granularities. While the model excels in anomaly localization on the MVTec dataset, its performance is slightly suboptimal on Bottle-Cap. These experiments provide quantitative evidence of the FEGAN' s robustness and adaptability for anomaly detection and localization tasks across diverse objects represented in multi-category datasets.

To provide a more intuitive understanding of anomaly localization, we present a qualitative discussion below. Fig.6 illustrates the anomaly localization results for 10 object classes in the MVTec. We selected two images from each class, featuring larger and smaller anomaly, respectively, for analysis. In the anomaly heat map, colors range from blue to red to indicate low to high anomaly likelihood. The green area in the anomaly segmentation map represents the model's localization area, while the red stroke delineates the actual anomaly boundary. As observed in the figure, FEGAN accurately identifies corresponding regions for anomaly of varying sizes in objects such as metal nuts, lenses, cables, toothbrushes, pills, etc. The heat map approximates the surface anomaly localization compared to real anomaly regions,
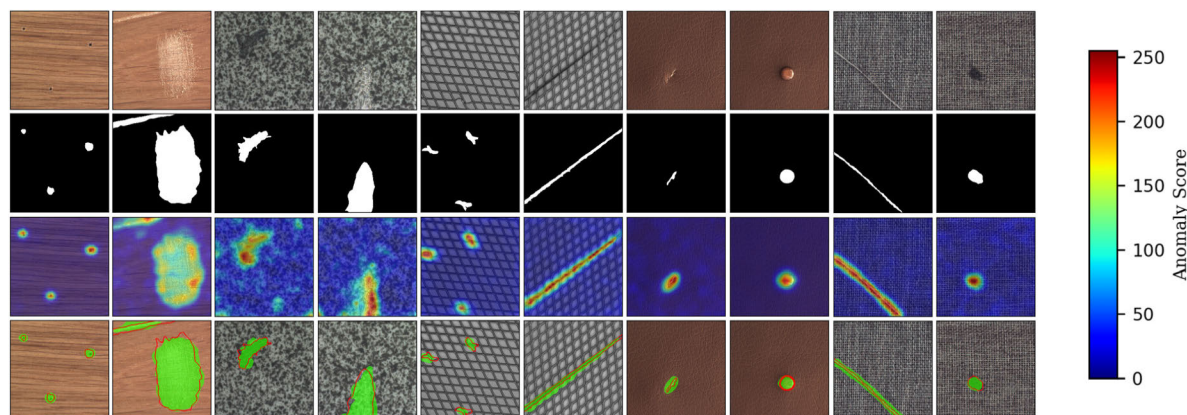
FIGURE 7. Anomaly localization results of FEGAN on texture classes of MVTec.The first line is the abnormal image. The second line is the ground truth. The third line shows the abnormal heat map. The last line gives the anomaly location map.

FIGURE 8. Partial anomaly localization results of the model on Bottle-Cap.The first line is the abnormal image. The second line is the ground truth. The third line shows the abnormal heat map. The last line gives the anomaly location map.
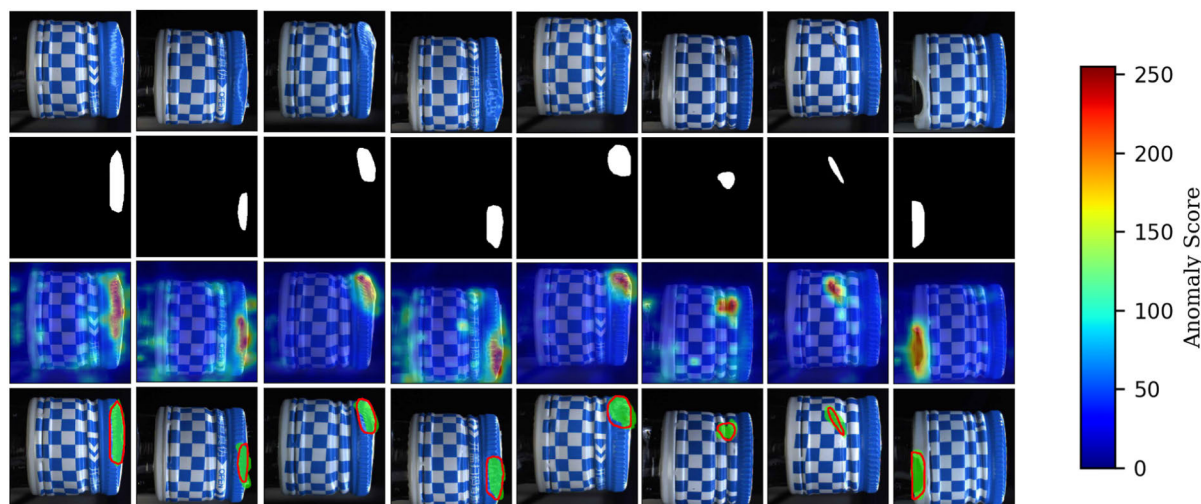
while the segmentation map precisely delineates the anomaly regions.

After extensive experimental validation, our model demonstrates strong localization capabilities. Fig. 7 showcases the model's localization outcomes for five distinct image types encompassing various texture features in MVTec. Notably, the fourth row presents finely detailed anomaly segmentation maps, offering clear delineation of the specific anomaly locations and their area ranges. The experiments reveal the model's robustness in identifying anomaly such as stains, scratches, and lines, showcasing a certain level of accuracy and indication in anomaly localization across different regions.

The qualitative discussion of the Bottle-Cap is unfolded below, and Fig. 8 visualizes some of the localization results. As depicted, the localization task aims to precisely pinpoint abnormal categories. The green areas denoting localization in the fourth row show significant alignment with the actual red boundaries, although some errors persist, indicating room for enhancement in localization performance.

After the above discussion on the visualization of abnormality segmentation, the specific location of the abnormality on the product surface is marked intuitively. By examining the abnormality segmentation map, one can effectively discern information regarding the size, shape, and quantity of abnormalities, aiding in their identification, classification, and positioning in the actual production line. Experimentation reveals that our method outputs highly coincident results with the actual anomaly locations and exhibits robustness across various detection targets. This capability offers a more accurate and reliable reference for quality inspection of industrial surface images in real-world scenarios.

D. RELATED DISCUSSION

In the preceding section, our proposed model demonstrates outstanding performance. However, upon comprehensive analysis of qualitative and quantitative results, we have identified some challenges that require attention and further investigation. These issues are outlined below:

i) FEGAN exhibits strong performance in certain categories but performs poorly in others. This discrepancy may stem from the algorithm's inability to generalize effectively across objects with varying appearances, textures, or shapes. Given the diverse feature distributions and anomaly patterns in the real world, greater adaptability and robustness are essential for successful anomaly detection.

ii) Our algorithm relies on key feature representations to differentiate between normal and abnormal samples. However, our analysis reveals that FEGAN's fine feature extraction in categories such as bottles, capsules, and screws is insufficient. In future research, we plan to explore effective feature fusion strategies to enhance the network's capability in representing features.

iii) Anomaly definitions may vary across different image categories. In our study, FEGAN employs a uniform anomaly metric, which may introduce limitations and ambiguities. To address this, we intend to tailor anomaly definitions for specific classes and utilize appropriate algorithm parameters to improve performance.

Furthermore, we need to consider the generalization ability of our method across diverse industrial environments. Our approach forces the model to adequately learn key features of normal patterns in images by training normal data. In order to improve the robustness of the network in practical applications, M-SE data enhancement technology is used to simulate various possible anomaly modes. We believe that the abnormal patterns in the images have a common feature of "mutation". In this work, we deliberately selected MVTec, a specialized dataset covering 15 categories, considering the complexity of anomaly detection of industrial products. In addition, we also selected the Bottle-Cap dataset, which is very common in the detection. After being verified on a wide range of classes, the proposed method shows optimal and sub-optimal performance in most classes. This shows that our method has a certain degree of multi-scenario adaptability and can maintain a high generalization ability across different product types. Nevertheless, we still support continuous learning and model updates. We will target new research objects and fine-tune the model parameters to adapt them to new application scenarios.

Lastly, it is important to acknowledge the computational resource and hardware limitations associated with our approach during actual operation. We conducted tests on devices with limited computing power, and FEGAN's inference speed was found to be slow under these conditions. However, on high-performance computing devices, our model can perform inference quickly and handle larger-scale image data effectively. Therefore, we recommend deploying FEGAN on high-performance GPU devices for optimal performance and scalability in real-world applications.

V. ABLATION STUDY

We conducted three ablation experiments to examine the influence of various factors on the model's performance. Firstly, we delved into the impact of the weighting parameters used in the method on both detection and localization. Secondly, we scrutinized the effectiveness of the M-SE. Lastly, we assessed the effects of different components and configurations on the experimental outcomes.

A. HYPERPARAMETERS

This section discusses the effects of different hyperparameters on the model performance to illustrate the scientific reliability of the weighting parameters. The hyperparameters in the loss function represent the importance of each component, and the calculation formula is shown in Eq. (5). This section discusses the values of the parameters from the range of 1 to 90, as shown in Fig. 9, with λ_{fea} and λ_{con} parameter values increase, the AUC-ROC of the model first increases sharply and then decreases gradually; as the λ_{adv} parameter values increase, the AUC-ROC of the model decreases steeply from about 0.95 to 0.63 and then rises oscillating. The experiment shows that the λ_{fea}, λ_{con} and λ_{adv} network performance is optimal when 30, 30, and 1 are taken, respectively.
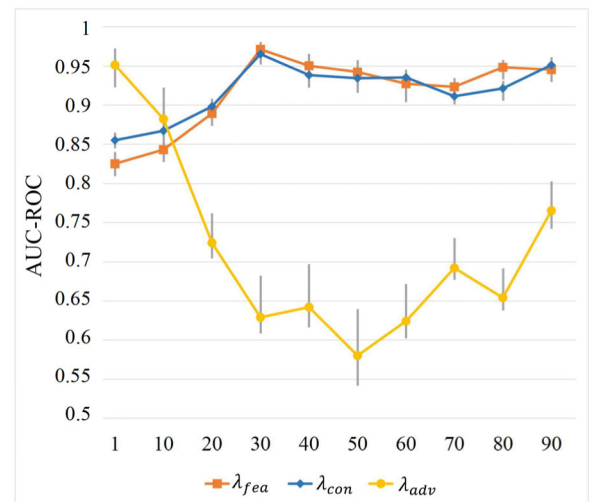

**FIGURE 9.** Network performance of the loss function hyperparameters under different values.

In addition, the parameters of the anomaly scores are equally critical to performance and are calculated as shown in Eq. (8). This section discusses the parameter λ values in the range of 0.1 to 0.9. As seen in Fig. 10, the difference between image-level and pixel-level detection performance is not very large, and the model localization performance is correlated with the detection performance. The experiment shows that when $\lambda = 0.6$, the detection and localization performance of the model is optimal for quantitative analysis.

B. ENHANCEMENT METHODS

M-SE plays a crucial role in our approach. Since the intensity of M-SE is controlled by the area ratio, the experimental effects of FEGAN under different enhancement conditions are discussed below. We examined the impact of different enhancement conditions on FEGAN's performance through

TABLE 5. Quantitative results of FEGAN under enhanced conditions with various area ratios. Data underlined and bolded in the table indicate optimal values.

Metircs	area=1%	area=5%	area=10%	area=20%	area=30%	area=50%	area=70%	area=100%
AUC-ROC (image-level)	0.921	0.944	0.965	**0.978**	0.967	0.942	0.901	0.873
AUC-ROC (pixel-level)	0.943	0.965	**0.985**	0.982	0.979	0.959	0.925	0.891
IOU	0.347	0.366	0.381	**0.399**	0.389	0.352	0.328	0.286
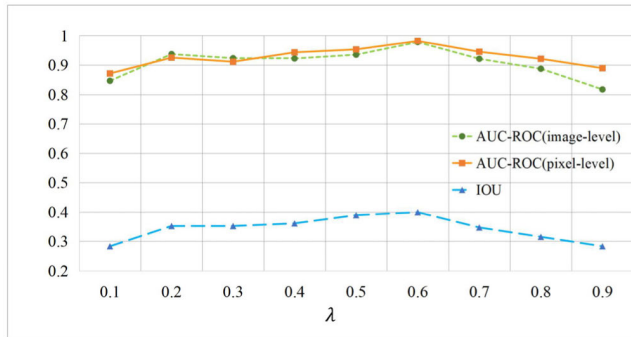
FIGURE 10. Effect of different weighting parameters on model detection and localization. Demonstrates the effect on image-level AUC-ROC, pixel-level AUC-ROC and IoU at different values.

experiments. Specifically, we conducted experiments with eight different area proportion ranging from 1% to 100%. In Table 5, area = 1% signifies that only one patch enhances the image, while area = 100% means that the entire image is sampled, resulting in extensive coverage by the enhanced patches. It's important to note that we allowed repeated overlays of multiple patches to avoid full coverage of the original image.

The results in Table 5 reveal that the network achieves its best image-level AUC-ROC and IoU values at area = 20% for M-SE. Although the optimal performance is not attained in pixel-level detection, its performance is very close to the results obtained with area = 10%. After thorough consideration, we selected area = 20% as the conditional attribute of M-SE. Additionally, when the AREA of M-SE is set to 1% or 100%, FEGAN' s detection and localization performance suffers. This indicates that both too small and too large an area intensity for the self-enhancement strategy are detrimental to our work. Too little intensity weakens the enhancement effect, while too much intensity masks the overall image features.Therefore, we should be careful about the choice of area strength.

To visualize the enhancement effect of M-SE at different area intensities, we have illustrated some images in Fig. 11. This figure showcases the impact of M-SE implementation on both texture and object classes. As depicted, the enhancement area expands as the percentage of copied area increases. Particularly, when area = 20%, the M-SE strategy notably enhances the model quality.

To assess the impact of various image enhancement strategies on the overall performance of the model, this section conducts experiments using different enhancement

TABLE 6. Effect of different enhancement strategies on experimental results. Data underlined and bolded in the table indicate optimal values.

Metircs	Cutout	Cutout (colour)	RIAD	SCADN	M-SE
AUC-ROC (image-level)	0.656	0.762	0.869	0.871	**0.978**
AUC-ROC (pixel-level)	0.691	0.747	0.883	0.892	**0.982**
IOU	0.232	0.279	0.339	0.362	**0.399**

methods. Table 6 presents the experimental results of five image enhancement techniques employed in the models discussed in this part. Cutout [46] randomly overlays only black pixel patches in the image. Cutout(colour) superimposes colors on top of the former. RIAD [27] divides the image into a grid of rectangular regions of size K × K pixels and non-intersecting each other, and forms an image mask by removing part of the pixel regions in the image. SCADN [25] achieves the removal of some regions from normal samples by generating striped masks with different scales in vertical and horizontal directions.

From Table 6, it's evident that Cutout and Cutout (colour) yield poor performance in surface anomaly detection and localization. RIAD contributes to improved model performance, achieving over 80% AUC-ROC at different grain sizes. SCADN achieves suboptimal performance in both detection and localization. Conversely, FEGAN enhanced by M-SE outperforms the aforementioned methods across all performance metrics, indicating that M-SE can accurately detect and localize anomaly, effectively enhancing the model's sensitivity.

In summary, it appears that M-SE effectively highlights simulated anomaly in the image. FEGAN notably enhances the visual effect and reconstruction quality of the image by learning specific anomalous properties. Through extensive training with enhanced images, FEGAN enhances the ability to identify and locate anomaly. In addition, M-SE strategy is simple and efficient, which is suitable for large-scale image training tasks. Experiments demonstrate the reliability of M-SE in this context.

C. MODEL COMPONENTS AND LOSS FUNCTION
In order to comprehensively assess the impact of each component in our proposed FEGAN model on its overall performance, we conducted experiments using various control variables to elucidate the importance of each component.
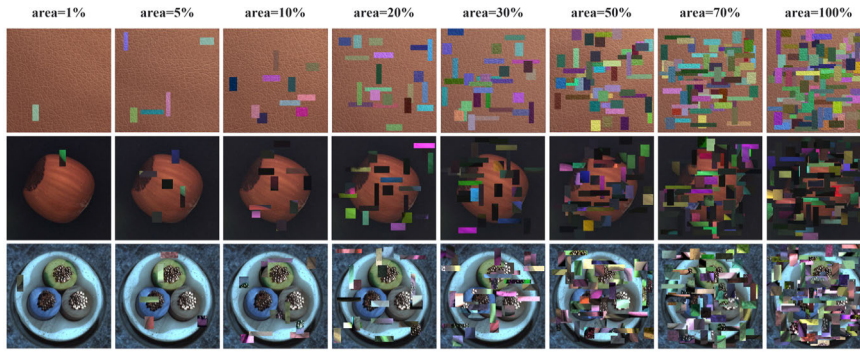
FIGURE 11. Enhancement effect of M-SE under various area ratio conditions.

TABLE 7. Quantitative results of FEGAN under enhanced conditions with various area ratios. Data underlined and bolded in the table indicate optimal values.

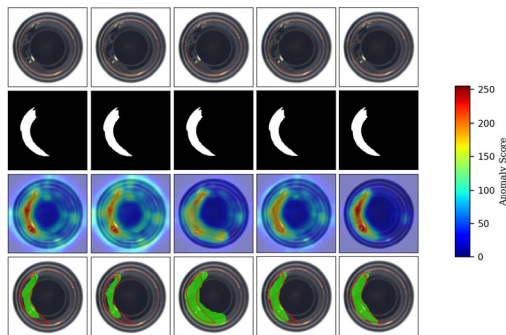| Generator	Discriminator	FEN	M-SE	Loss	AUC-ROC (image-level)	AUC-ROC (pixel-level)	IoU
√				$\mathcal{L}_{con}$	0.623	0.637	0.215
√			√	$\mathcal{L}_{con}$	0.782	0.757	0.254
√	√		√	$\mathcal{L}_{con}\ \mathcal{L}_{adv}$	0.804	0.815	0.328
√		√	√	$\mathcal{L}_{con}\ \mathcal{L}_{fea}$	0.927	0.953	0.374
√	√	√	√	$\mathcal{L}_{con}\ \mathcal{L}_{adv}\ \mathcal{L}_{fea}$	**<u>0.978</u>**	**<u>0.982</u>**	**<u>0.399</u>**

FIGURE 12. Visualization results under different experimental conditions. The first row is the lens image having the same anomaly, the second row is the binarized image of the real anomaly, the third row is the anomaly detection heat map, and the fourth row is the anomaly segmentation image.

As depicted in Table 7, the first row represents the baseline model, utilizing only the generator and content loss for training. The results are relatively unsatisfactory, reflecting the limited feature extraction capability and generalization of the model. In the second row, the addition of M-SE significantly enhances the model's performance. Moving to the third row, the inclusion of a discriminator along with discriminative loss further improves the model's feature learning ability through adversarial training, resulting in a pixel-level AUC-ROC of 0.815. In the fourth row, FEN is integrated to extract deep image features based on the baseline model with image enhancement. While this model achieves basic anomaly classification and pixel localization, its accuracy requires further enhancement. The last row incorporates all aforementioned components with content loss, adversarial loss, and feature loss as optimization objectives. Experimental results indicate that our model exhibits high accuracy, with each component significantly contributing to its overall performance.

To visually demonstrate the impact of anomaly localization under various model structures, we selected identical anomaly images to illustrate the five cases outlined in Table 6. Moving from left to right in Fig. 10, the anomaly prediction region of the model progressively approaches the red boundary of the real anomaly. This sequential depiction effectively illustrates how each component contributes positively to the model's prediction accuracy, leading to gradual improvement.

VI. CONCLUSION
In this paper, we propose FEGAN, a novel anomaly detection and localization method for surface anomaly in industrial products. We have built a feature extraction network (FEN) and an improved generative adversarial network (IGAN). By combining deep features and image features, FEGAN detects anomaly effectively. Additionally, we propose a multi-scale image self-enhancement (M-SE) method to enhance the model's robustness and accuracy. Experimental results demonstrate that our method achieves optimal AUC performance on both the MVTec and Bottle-Cap datasets, while also exhibiting reliable anomaly localization capabilities.

In our future research, we aim to explore two key areas. Firstly, we plan to enhance anomaly localization by guiding the model to focus specifically on anomalous regions in the image. This involves designing and implementing an anomaly attention mechanism to further refine the segmentation of anomaly. Secondly, we aim to improve the efficiency of real-time anomaly detection and localization for industrial products. Additionally, we will focus on expanding the practical application scenarios of our method to address the complex needs of industrial production.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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. 1–11, 2021.

[7] 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. 2019, pp. 40–48, Jun. 2019.

[8] J. Peng, H. Shao, Y. Xiao, B. Cai, and B. Liu, "Industrial surface defect detection and localization using multi-scale information focusing and enhancement GANomaly," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122361.

[9] W. Khan and M. Haroon, "An unsupervised deep learning ensemble model for anomaly detection in static attributed social networks," *Int. J. Cognit. Comput. Eng.*, vol. 3, pp. 153–160, Jun. 2022.

[10] W. Khan, S. Abidin, M. Arif, M. Ishrat, M. Haleem, A. A. Shaikh, N. A. Farooqui, and S. M. Faisal, "Anomalous node detection in attributed social networks using dual variational autoencoder with generative adversarial networks," *Data Sci. Manage.*, vol. 7, no. 2, pp. 89–98, Jun. 2024.

[11] W. Khan and M. Haroon, "A pilot study and survey on methods for anomaly detection in online social networks," in *Human-Centric Smart Computing*. Cham, Switzerland: Springer, 2022, pp. 119–128.

[12] H. Nizam, S. Zafar, Z. Lv, F. Wang, and X. Hu, "Real-time deep anomaly detection framework for multivariate time-series data in industrial IoT," *IEEE Sensors J.*, vol. 22, no. 23, pp. 22836–22849, Dec. 2022.

[13] Z. Ji, Y. Wang, K. Yan, X. Xie, Y. Xiang, and J. Huang, "A space-embedding strategy for anomaly detection in multivariate time series," *Expert Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117892.

[14] M. Hu, X. Feng, Z. Ji, K. Yan, and S. Zhou, "A novel computational approach for discord search with local recurrence rates in multivariate time series," *Inf. Sci.*, vol. 477, pp. 220–233, Mar. 2019.

[15] X. Deng, L. Xiao, X. Liu, and X. Zhang, "One-dimensional residual GANomaly network-based deep feature extraction model for complex industrial system fault detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. NIPS*, vol. 27, 2014, pp. 2672–2680.

[17] T. Schlegl, P. Seebock, 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.*, vol. 2017, pp. 146–157.

[18] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. Van Den Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1705–1714.

[19] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," 2018, *arXiv:1807.02011*.

[20] 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.

[21] G. Zhang, K. Cui, T.-Y. Hung, and S. Lu, "Defect-GAN: High-fidelity defect synthesis for automated defect inspection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2523–2533.

[22] V. Zavrtanik, M. Kristan, and D. Skočaj, "Draem—A discriminatively trained reconstruction embedding for surface anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vision.*, 2021, pp. 8330–8339.

[23] S. Niu, B. Li, X. Wang, and Y. Peng, "Region- and strength-controllable GAN for defect generation and segmentation in industrial images," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4531–4541, Jul. 2022.

[24] Y. Gao, L. Gao, and X. Li, "A generative adversarial network based deep learning method for low-quality defect image reconstruction and recognition," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3231–3240, May 2021.

[25] 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, no. 4, pp. 3110–3118.

[26] M. Zaigham 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.

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

[28] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3379–3388.

[29] 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.*, Jul. 2019, pp. 1–8.

[30] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: A patch distribution modelling framework for anomaly detection and localisation," in *Proc. Int. Conf. Pattern Recognit.*, 2021, pp. 475–489.

[31] M. Yang, P. Wu, and H. Feng, "MemSeg: A semi-supervised method for image surface defect detection using differences and commonalities," *Eng. Appl. Artif. Intell.*, vol. 119, Jul. 2023, Art. no. 105835.

[32] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," 2020, *arXiv:2005.02357*.

[33] J. Yi and S. Yoon, "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Nov. 2020, pp. 375–390.

[34] 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.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[36] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, Aug. 2015.

[37] J. Yang, Y. Shi, and Z. Qi, "Learning deep feature correspondence for unsupervised anomaly detection and segmentation," *Pattern Recognition*, vol. 132, Dec. 2022, Art. no. 108874.

[38] Y. Dai, L. Zhang, F.-Y. Fan, Y.-J. Wu, and Z.-K. Zhao, "SCGAN: Extract features from normal semantics for unsupervised anomaly detection," *IEEE Access*, vol. 11, pp. 137957–137968, 2023.

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

[40] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1–12.
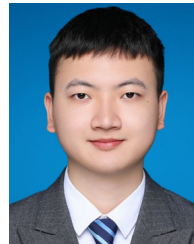
[41] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "GANomaly: Semi-supervised anomaly detection via adversarial training," in *Computer Vision—ACCV*. Cham, Switzerland: Springer, 2018, pp. 622–637.

[42] C. Huang, J. Cao, F. Ye, M. Li, Y. Zhang, and C. Lu, "Inverse-transform autoencoder for anomaly detection," 2019, *arXiv:1911.10676*.

[43] 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.

[44] Y. Shi, J. Yang, and Z. Qi, "Unsupervised anomaly segmentation via deep feature reconstruction," *Neurocomputing*, vol. 424, pp. 9–22, Feb. 2021.

[45] D. Dehaene, O. Frigo, S. Combrexelle, and P. Eline, "Iterative energy-based projection on a normal data manifold for anomaly localisation," 2020, *arXiv:2002.03734*.

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

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 Sichuan Digital Economy Research Institute. From 2021 to 2024, she has published a total of five papers. 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 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 papers. His research interests include artificial intelligence and quantum computing. He is a member of CCF (18135M).

YANG DAI was born in Chengdu, China, in 1997. He received the master's degree in electronic information from China West Normal University, in 2024. He is currently a Scientific Researcher with Sichuan Digital Economy Research Institute. From 2021 to 2024, he has published a total of five papers. His research interests include computer vision and deep learning.

• • •