Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on L-PBF Defect Detection Based on VAE-GAN Data Augmentation with Multi-Head Attention Mechanism

Yanhui Ma
Lusheng Li
Chengke Wang
Shuo Li
Bin Fu
Zhiqiong Wang

zqwang315@tju.edu.cn

Tianjin University of Technology https://orcid.org/0000-0003-1430-0795

Research Article

Keywords: Additive Manufacturing, Data Augmentation, Small-Sample Learning, Defect Detection

Posted Date: August 11th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7036982/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Research on L-PBF Defect Detection Based on VAE-GAN Data Augmentation with Multi-Head Attention Mechanism

Yanhui Ma, Lusheng Li,Chengke Wang,Shuo Li,Bin Fu and Zhiqiong Wang

School of Management, Tianjin University of Technology, TianJin 300384, China

E-mail: zqwang315@tju.edu.cn

Abstract

Additive Manufacturing (AM), particularly Laser Powder Bed Fusion (L-PBF), faces critical challenges in defect detection due to the scarcity of high-quality training data and severe class imbalance, which significantly degrade the accuracy of deep learning models. To address these issues, this study proposes a novel data augmentation framework combining geometric transformations with an enhanced Variational Autoencoder-Generative Adversarial Network (VAE-GAN). Traditional augmentation techniques (rotation, scaling, flipping) are first applied to alleviate sample imbalance, followed by the improved VAE-GAN to synthesize high-fidelity defect images, thereby enriching dataset diversity. Experimental results on an L-PBF defect dataset demonstrate significant improvements in detection performance: This study conducted a comparative experiment evaluating the performance of YOLOv4, YOLOv7, YOLOv8, SSD, and Faster R-CNN on defect detection tasks before and after data augmentation. The results demonstrated significant mAP improvements across all models, with YOLOv4 achieving the most substantial enhancement (+20.03%, from 65.18% to 85.21%) despite its lower baseline performance. Faster R-CNN attained the highest post-augmentation mAP (87.75%), representing the best overall performance. YOLOv8 exhibited an optimal balance between real-time processing and accuracy (67.07%→84.66%, +17.59%), approaching Faster R-CNN's performance level. While SSD showed the smallest improvement (+12.34%), it maintained a relatively high baseline mAP (73.99%). These results validate the effectiveness of the proposed method in overcoming data scarcity and improving defect detection accuracy in AM.

Keywords: Additive Manufacturing, Data Augmentation, Small-Sample Learning, Defect Detection

1. Introduction

Additive Manufacturing (AM) is an advanced manufacturing technology that has been widely adopted in industries such as aerospace, medical, and automotive. Its advantages, including the ability to fabricate complex geometric structures, high material utilization, and customized production, make it a preferred choice[1]. However, the complexity and instability of the AM process often lead to defects such as porosity, lack of fusion, and cracks, which significantly affect the mechanical properties and service life of the final product [2]. Among these, porosity and cracks are the most common and impactful defects. Therefore, developing efficient and accurate defect detection methods is of great significance to ensure the structural integrity and performance of AM components.

Traditional defect detection methods can be categorized into offline and online approaches. Offline methods, such as computed tomography (CT) and ultrasonic testing, offer high precision but are costly and time-consuming. Online methods, including optical thermography and high-speed X-ray imaging, enable real-time monitoring but suffer from low detection rates and poor environmental adaptability. Moreover, these methods typically involve high costs and complex data processing procedures [3][4]

With the rapid advancement of deep learning, image-based defect detection methods have gained considerable attention. Convolutional Neural Networks (CNNs) have been employed for defect detection in AM, improving detection accuracy [5]. However, the performance of CNN models heavily depends on the quality and quantity of the training data. Since defect data collection is expensive, the imbalance between positive and negative samples under small-sample conditions makes it challenging for models to effectively learn defect features.

Data augmentation techniques provide an effective solution to address data scarcity and imbalance. Traditional data augmentation methods, such as geometric transformations (rotation, scaling, flipping) and color adjustments (brightness and contrast enhancement), increase data diversity but fail to generate entirely new samples, thus limiting the expansion of the sample distribution [6]. To tackle the small-sample problem, deep generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have demonstrated great potential in data augmentation. GANs utilize adversarial training between a generator and a discriminator to learn the true data distribution and generate high-quality samples. Researchers have leveraged GANs to synthesize AM defect images to mitigate data imbalance issues. For instance, Jihoon Chung et al.[7]combined GANs with classifier optimization strategies to enhance the stability and accuracy of anomaly detection.

However, conventional GANs suffer from mode collapse and unstable training, while VAEs, although capable of generating images, tend to produce blurry outputs. The VAE-GAN[8][9] model integrates the advantages of both approaches: VAE employs variational inference to learn the latent distribution of data, mitigating the mode collapse issue in GANs, while the adversarial training of GAN ensures the generated images maintain high quality and diversity, thereby achieving a balance between global and local features. Yitian Wang et al.[10]successfully applied VAE-GAN to augment a wafer dataset, improving classifier detection accuracy and validating the effectiveness of this approach in small-sample data augmentation.

Despite the promising performance of VAE-GAN in image generation tasks, it may struggle to capture fine-grained details in complex defect images, leading to suboptimal image quality and diversity. To address this issue, we introduce a Multi-Head Attention (MHA) mechanism into VAE-GAN, enabling parallel computation across multiple attention heads to learn feature representations from different subspaces. This enhancement improves the model's capability to capture detailed image features, thereby further enhancing the quality and diversity of generated images.

This study focuses on the detection of porosity and crack defects, which are the most prevalent and quality-critical defect types in AM. First, we employ conventional data augmentation techniques, such as rotation, scaling, flipping, and cropping, to expand the existing defect dataset and alleviate data imbalance. Subsequently, we utilize the improved VAE-GAN to generate high-quality synthetic defect images, further increasing sample diversity and improving the generalization ability of the detection model. Finally, we construct a defect detection model and conduct experimental validation on datasets before and after augmentation to evaluate the effectiveness of the proposed approach.

The innovations of this study include: (1) Hierarchical Augmentation Framework Integrating Geometric and Generative Models We propose a novel hierarchical framework that synergistically combines geometric transformations (e.g., rotation, scaling) with an improved VAE-GAN architecture. Unlike conventional methods that apply these techniques in isolation, our framework leverages geometric transformations to preserve low-level defect features (e.g., edge continuity) while utilizing the generative model to synthesize high-fidelity defect patterns. This dual-stage approach mitigates the limitations of traditional augmentation in generating novel defect morphologies.(2) To overcome the blurred outputs and insufficient feature diversity of standard VAE-GANs, we introduce a Multi-Head Attention (MHA) mechanism into both the encoder and generator. The MHA module enables parallel computation across multiple attention heads, dynamically focusing on critical defect regions (e.g., crack tips, pore boundaries) at different scales. This design enhances the model's capability to capture fine-grained details while maintaining global structural consistency.(3) By introducing a scaled latent sampling perturbation mechanism in the VAE-GAN framework—namely, using low-variance Gaussian noise in the reparameterization process—we effectively control the perturbation intensity in the latent space. This design significantly enhances the model's noise robustness and training stability, leading to faster convergence in the early training stages.

The remainder of this paper is organized as follows: Chapter 2 introduces related work, including data augmentation methods, defect detection methods, and attention mechanisms; Chapter 3 details the proposed data augmentation algorithm and defect detection model; Chapter 4 validates the effectiveness of the proposed method through experiments; and Chapter 5 summarizes the research findings and outlines future research directions.

2.Related work

2.1. Traditional Data Augmentation Methods

Data augmentation methods primarily include traditional methods and intelligent methods based on generative models[11]. Traditional data augmentation techniques mainly involve linear transformations to slightly modify image data, preserving the original label information while increasing the data volume. This improves dataset quality and mitigates model overfitting[12][13]. These method s primarily include operations such as geometric transformations, color transformations, and noise transformations. For instance, Perez et al. [14] demonstrated that these methods primarily alter low-level image statistics

(e.g., pixel positions) without creating novel defect patterns, leading to limited generalization in complex scenarios such as multi-scale porosity clusters. Matsunaga et al.[15] validated the effectiveness of test-time augmentation in skin lesion classification tasks using geometric transformations such as rotation, translation, scaling, and flipping.

In the field of additive manufacturing, Wenyuan Cui *et al*.[16] expanded their dataset by applying traditional data augmentation techniques such as random rotation, flipping, and cropping to optical images of both normal and defective samples. Using a convolutional neural network (CNN) for defect detection, their method achieved an accuracy of 92.1%. Additionally, Choi *et al*.[17] verified the effectiveness of object detection methods in additive manufacturing by augmenting defect images through transformations such as rotation, downscaling, and shifting.

Although these methods can enhance data diversity and quantity while partially mitigating data imbalance or limited sample sizes issues, their effectiveness in defect detection remains constrained—particularly in scenarios with limited sample sizes or severe class imbalance. Since traditional augmentation techniques primarily modify existing image structures through geometric or photometric transformations without generating novel defect patterns, they may fail to substantially improve the model's capability to recognize complex or rare defects. Consequently, the performance gains in defect detection are often marginal, highlighting the need for more advanced augmentation strategies tailored to address data scarcity and imbalance.

2.2. Deep Generative Models

2.2.1 VAE/GAN Fundamental Theories

Deep generative models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) have emerged as two popular deep generative models widely applied in image generation[18], image transformation[19], image resolution enhancement[20], domain adaptation[21], and anomal y detection[21]. For instance, Jang Young In *et al.*[22] utilized VAEs to reconstruct synthetic electrocardiogram (ECG) signals, thereby improving the consistency of the standard deviation of normal-to-normal intervals (SDNN).

GANs, introduced by Goodfellow in 2014, represent a remarkable class of unsupervised generative neural networks[23]. Designed as a minimax game between two subnetworks—namely, the generator and discriminator—GANs learn to approximate the data distribution[24].

Numerous researchers have conducted extensive studies on the application of Generative Adversarial Networks (GANs) for augmenting limited fault sample data. For instance, Zhang *et al.*[25] employed GANs to enhance wind turbine vibration signals, addressing data imbalance challenges. Yang *et al.*[26] used Conditional GANs (CGANs) to learn the real distribution of bearing fault samples, expanding the dataset and improving fault diagnosis accuracy using convolutional neural networks (CNNs). Ngoc-Trung Tran *et al.*[27] proposed a GAN -based data augmentation optimization model and validated its superior Fréchet Inception Distance (FID) performance on natural and medical datasets. Frid-Adar *et al.*[28] applied GAN-based image synthesis techniques for liver lesion classification, significantly improving sensitivity and specificity. Furthermore, Hongbin Gao *et al*.[29] combined Deep Convolutional GANs (DCGANs) with traditional data augmentation methods to increase training data volume, with plans to employ a lightweight CNN model based on VGG11(Visual Geometry Group 11-layer) for defect classification.

While both are generative models, their fundamental differences lie in Generative Adversarial Networks (GANs) offer significant advantages in image generation, they still face several challenges in practical applications, such as training instability, mode collapse, gradient vanishing, and weak controllability of generated data[30]. Unlike GANs, Variational Autoencoders (VAEs) optimize the learning process through variational inference[31] and generate samples by sampling from a random noise distribution. However, due to the lack of an adversarial learning mechanism in the reconstruction process, the generated images often suffer from blurriness. Additionally, the optimization of VAE models relies on manually designed parameters, which may introduce unavoidable noise and degrade the quality of the generated samples.

2.2.2 Applications in AM

In additive manufacturing research, the VAE model is predominantly employed as a feature extractor. For instance, Zihan Wang & Hongyi Xu[32] utilized a Variational Autoencoder (VAE) as a feature extractor to map 3D metamaterial geometries into a low-dimensional latent feature space. This latent feature space was simultaneously linked to a discriminator/regressor to predict manufacturability metrics and mechanical properties. William Frieden Templeton *et al*.[33] mentioned a VAE regression model to extract spatial features, predict fatigue life, and identify pore defect characteristics governing fatigue behavior. Ertay, D. S. *et al.*[34] obtained a secondary dataset containing the pore space of fabricated components via X-ray computed tomography (CT) and registered it into a synthetic dataset. Machine learning models, namely a Conditional Variational Autoencoder (CVAE) and a Convolutional Neural Network (CNN), were

then trained based on input features to predict pore formation.

In the field of additive manufacturing, numerous studies have leveraged GANs and their variants to generate high-quality images, thereby addressing data imbalance issues.For example, Gobert *et al.*[35] employed Conditional GANs (cGANs) to generate layer-wise images from the metal additive manufacturing process. Jihoon Chung *et al.*[36] optimized a standard GAN alongside a classifier to stabilize the learning process and applied it to anomaly detection in real-world additive manufacturing scenarios. Similarly, Zhibo Zhang *et al.*[37] utilized machine learning techniques to predict melt pool size and morphology, synthesizing melt pool images via a Melt Pool GAN (MP-GAN) that achieved a structural similarity index (SSIM) of 0.91.Gobert *et al.* [38] employed conditional GANs to generate layer-wise images from metal additive manufacturing processes. Jihoon Chung et al. developed an optimized framework integrating standard GANs with classifier techniques for anomaly detection in practical additive manufacturing, which successfully identified process anomalies even in highly imbalanced datasets.

2.3. Hybrid VAE-GAN Frameworks

To address the respective limitations of GANs and VAEs, researchers have proposed the VAE-GAN model, which combines the representation learning capability of VAEs with the adversarial generation ability of GANs. In VAE-GAN, the VAE is responsible for learning latent representations of the data and generating diverse images, while the GAN discriminator optimizes the generated samples to enhance image clarity and realism. Yitian Wang *et al.*[10] leveraged the advantages of VAE-GAN to achieve high-quality data augmentation and improve wafer classification accuracy. Similarly, Sun *et al.*[39] proposed a VAE-GAN-based energy theft detection model to address the issue of data imbalance in energy theft detection. Their model generated synthetic data that accurately captured real energy theft characteristics, which were then used to train the detector, effectively mitigating the impact of data imbalance. In the field of additive manufacturing, Zheren Song *et al.*[40] proposed a novel hybrid deep generative prediction network (HDGPN) that leverages both variational autoencoder and generative adversarial network is proposed to characterize the complex pore microstructure with in-depth representations and predict pore morphology under arbitrary processing parameters. By visualizing the predicted pore morphology, the complicated interaction dynamics between the processing parameters and pore microstructure are directly revealed, which may guide the optimization of metal AM manufacturing processes to fabricate defect-free products.

Although the VAE-GAN model combines the strengths of Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), it still has certain limitations. For instance, the generated images may lack fine-grained details, and the model's ability to capture complex dependencies in the data can be limited. These shortcomings arise because the VAE-GAN framework does not explicitly model the relationships between different regions of an image, which can lead to suboptimal feature representation and generation quality. To address these issues, researchers have turned to attention mechanisms, which have shown remarkable success in enhancing model performance across various computer vision tasks.

2.4. Attention Mechanisms

2.4.1 Fundamental Theory

The attention mechanism was first introduced by Bahdanau *et al.* in 2014 for neural machine translation (NMT). Its core idea is to dynamically assign different weights to different parts of the input sequence, thereby enhancing the model's ability to focus on critical information. This approach effectively mitigates the limitations of traditional sequential models, such as recurrent neural networks (RNNs), which struggle to capture long-range dependencies[41].

In 2017, Vaswani *et al*. proposed the multi-head attention mechanism (MHA) as a fundamental component of the Transformer architecture. MHA enhances information capture and improves generalization by enabling multiple independent attention heads to compute in parallel, allowing the model to learn diverse feature representations across different subspaces. The computation of MHA consists of the following steps: (1) The input is first linearly projected into query (Q), key (K), and value (V) matrices. (2) The scaled dot-product attention mechanism computes the similarity between queries and keys, followed by Softmax normalization to obtain attention weights. (3) The attention weights are then used to perform a weighted sum over the values, extracting the most relevant information. (4) The outputs from multiple attention heads are concatenated and projected through a linear transformation to produce the final output[42].

Compared to single-head attention, MHA enables the model to learn more diverse feature representations, leading to improved performance in long-range dependency modeling and complex learning tasks. As a result, the Transformer model and its variants—such as BERT, GPT, and Vision Transformer (ViT)—have achieved remarkable success in both natural language processing (NLP) and computer vision (CV) applications.

2.4.2 Applications in Detection

Attention mechanisms and multi-head attention mechanisms have been widely applied in defect detection and object detection research. For instance, in the application of attention mechanisms, Wang *et al.*[43] proposed Non-local Neural Networks, leveraging self-attention to model long-range dependencies, thereby improving performance in video classification and object detection. Chen *et al*.[44] introduced Vision Transformer (ViT), which partitions images into fixed-size patches and treats them as sequential data, applying Transformer-based architectures for image classification. ViT has demonstrated performance comparable to convolutional neural networks (CNNs). Zhang *et al.*[45] proposed a self-attention-based defect detection model that focuses on defect regions, enabling precise detection of minute defects, especially in complex backgrounds.

In the application of multi-head attention, An *et al.*[46] proposed Repulsive Attention, which introduces a repulsion term to encourage diversity among attention heads. Zhang *et al*.[47] introduced the Mixture of Attention Heads mechanism, dynamically selecting the most suitable attention head for each input token, thereby improving computational efficiency and representational power. Li *et al*.[48] proposed a defect detection method incorporating multi-scale attention mechanisms, leveraging multi-head self-attention to enhance the recognition of defects at different scales. Wang *et al.*[49] introduced a defect detection model based on Vision Transformer, integrating the global information modeling capability of self-attention with the local feature extraction ability of convolutional networks, significantly improving the accuracy of complex defect recognition. Zhou *et al*.[50] proposed a hybrid model that combines multi-head self-attention mechanisms with convolutional neural networks. By optimizing feature extraction through attention mechanisms, this model enhances both the accuracy and efficiency of defect detection, making it particularly suitable for high-noise environments.

In this paper, the multi-head attention mechanism is introduced into the VAE-GAN model to enhance its capability of capturing image details and improve the quality of generated images. By computing multiple attention heads in parallel, the multi-head attention mechanism can capture multi-level global and local features, addressing the limitations of VAE-GAN in modeling complex dependencies. This optimization of feature representation and generation further enhances the performance of data augmentation and detection tasks.

2.5. AM-specific Considerations

In additive manufacturing (AM), porosity and cracks are recognized as the two most prevalent process-induced defects, with formation mechanisms closely associated with material properties, processing parameters, and thermodynamic behavior . Porosity primarily originates from melt pool instability (e.g., lack-of-fusion pores caused by insufficient laser power or excessive scanning speed) or gas entrapment in raw materials (*e.g*., moisture absorption in powder feedstock or shielding gas incorporation into the melt pool), typically manifesting as spherical or irregularly-shaped voids [51][52].

Extensive research has revealed that pore size distribution ranges from micrometer-scale (<10 μm) to millimeter-scale, with subcritical pores serving as preferential sites for fatigue crack initiation under cyclic loading [53]. The spatial distribution of porosity exhibits significant anisotropy, with pores tending to accumulate along melt pool boundaries or interlayer interfaces, thereby forming interconnected defect networks [54]. Furthermore, near-surface pores can substantially reduce component service life by accelerating oxidation and corrosion processes [55].

Cracks predominantly result from residual stress accumulation induced by rapid cooling, particularly in high-temperature alloys or brittle materials. These defects can be categorized into solidification cracks (caused by intergranular liquid film rupture) and interlayer thermal stress cracks (resulting from temperature gradient variations between deposited layers) . Based on their formation stages, cracks are further classified into solidification cracks (associated with low-melting-point eutectic phases) and solid-state cracks (initiated by stress concentration at ductile-to-brittle transition temperatures). Crack propagation typically follows columnar grain boundaries or fusion lines, ultimately forming through-thickness fracture channels [56][57]. Notably, crack orientation demonstrates a distinct angular relationship with deposition paths, leading to anisotropic mechanical properties in the fabricated components [58].

3. Methodology

The entire process of defect detection in the powder bed fusion (PBF) process based on data augmentation is illustrated in **Figure 1**. The workflow consists of three main steps:

（1） Data Preprocessing: This is the first step in the defect detection process, aimed at cleaning and standardizing the raw data while balancing the number of defective samples to prevent imbalanced data from affecting the detection results. This step lays the foundation for subsequent analysis and detection. Let Data_original represent the raw dataset, which undergoes several operations, including rotation, brightness adjustment, contrast enhancement, saturation adjustment, and hue modification.

These operations ensure data standardization and consistency while addressing sample imbalance issues, thereby providing high-quality input data for subsequent modeling and detection.

（2）Data Augmentation: After preprocessing, the data is fed into the Multi-VAE-GAN model for advanced augmentation. Unlike traditional data augmentation methods, Multi-VAE-GAN generates realistic and diverse defect samples to expand the dataset. Through the enhancement of Multi-VAE-GAN, the dataset is enriched and optimized, offering an effective solution to the challenge of small-sample data imbalance.

（3）Defect Detection: Various object detection models are employed to train and evaluate the data, ultimately achieving accurate and efficient defect identification.

3.1. Data Augmentation Algorithms

3.1.1 Traditional Data Augmentation

Traditional data augmentation methods, such as geometric transformations, color adjustments, and pixel-wise modifications, have demonstrated significant effectiveness in expanding datasets and enhancing model generalization. However, their limitations have become increasingly evident. First, these methods rely on manually designed rules, making it difficult to capture complex structures and semantic information within the data. Second, traditional augmentation techniques typically generate only limited, low-level image transformations and cannot produce entirely new or diverse data samples. Consequently, as deep learning models become more complex, traditional data augmentation methods can no longer meet the demand for high-quality and diverse data.

3.1.2 Variational Autoencoder (VAE) Model

The Variational Autoencoder (VAE) [51] is a generative model based on probabilistic graphical models that learns latent representations of data for generating new samples. The fundamental idea of VAE is to model the probability distribution of the data, thereby enabling the generation of new samples that closely resemble the original data distribution. Unlike conventional autoencoders, VAE imposes a probabilistic structure on the latent space, ensuring both diversity and continuity in the generated data.

(1)Variational Autoencoder (VAE) Architecture and Workflow

The Variational Autoencoder (VAE) is a generative model composed of three core components: the encoder (E), latent variable sampling (*z*), and the decoder (D) [51]: Its workflow, as illustrated in **Figure 2**, can be summarized as follows: First, the input image is processed by the encoder E, which extracts latent features and maps them to the latent variable Z_E.Next, latent representations are sampled from the learned distribution using the reparameterization trick, generating the latent representation Z_E.Finally, the latent representation Z_E is fed into the generator *G*, which decodes it into a high-dimensional image for reconstruction.

(2) VAE Model Loss Function [51]

The objective of optimizing the VAE model is to learn the model parameters by maximizing the log-likelihood of the data, denoted as $\log p(x)$. However, directly computing $\log p(x)$ is challenging. To address this, VAE employs variational inference and introduces the Evidence Lower Bound (ELBO) as the optimization target. The ELBO consists of two main components:

1)Kullback-Leibler (KL) Divergence: The *KL* divergence measures the difference between two probability distributions. In VAE, it is used to quantify the discrepancy between the approximate posterior distribution q(z|x) and the prior distribution p(z). For two Gaussian distributions $q(z \mid x) = N(\mu_q, \sigma_q^2)$ and $p(z) = N(0,1)$, the *KL* divergence is given by:

$$KL(q(z \mid x) \| p(z)) = -\frac{1}{2}\sum_{i=1}^{d}(1 + \log(\sigma_{q,i}^2) - \mu_{q,i}^2 - \sigma_{q,i}^2) \quad (1)$$

where d is the dimensionality of the latent variable *z*, μ_q and σ_q^2 are the mean and variance of *q(z / x)*, respectively.

2)Reconstruction Loss: The reconstruction loss quantifies the difference between the reconstructed data x' and the original data x .Common reconstruction loss functions include Mean Squared Error (MSE) and Cross-Entropy Loss. For image data, MSE is typically used as the reconstruction loss, and its formulation is:

$$MSE(x, x') = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \quad (2)$$

Where n is the dimensionality of the data, and x and x' represent the i -th elements of the original and reconstructed data, respectively.

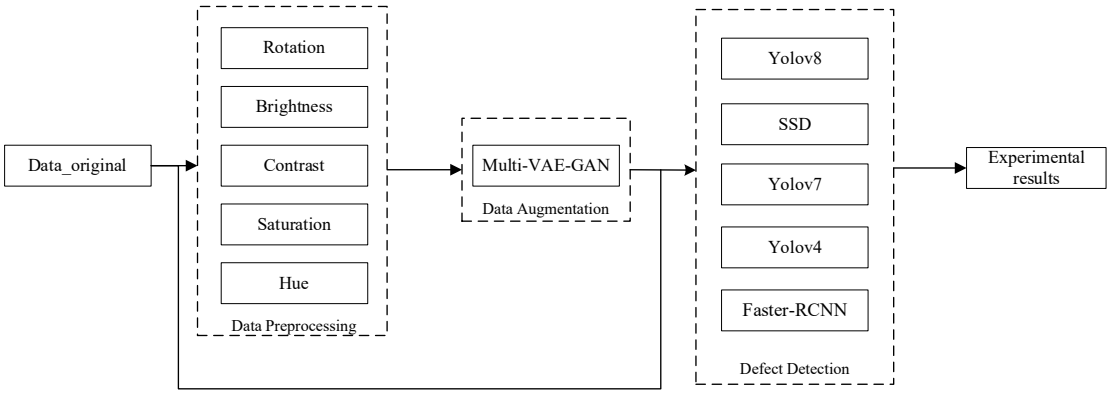

Figure 1. Overall Workflow of the Defect Detection Model

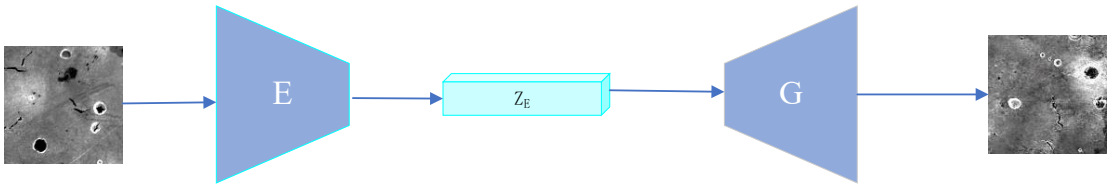

Figure 2. VAE Workflow Diagram

3.1.3 Generative Adversarial Network (GAN) Model

Generative Adversarial Networks (GANs) introduce an adversarial learning framework that significantly improves the quality of generated images. A GAN consists of two core components: the Generator (G) and the Discriminator (D). The generator synthesizes data from random noise, while the discriminator evaluates whether the input data is real (from the original dataset) or generated by the model. Through an adversarial learning process, the discriminator provides feedback to guide the generator's improvement [60].

(1) Core Concept and Workflow of the Generative Adversarial Network (GAN)

The core idea of the Generative Adversarial Network (GAN) is to optimize the model through adversarial learning between the generator (G) and the discriminator (D). The GAN model consists of two primary components: the generator and the discriminator. The generator produces synthetic samples $G(z)$ from random noise z, where z is typically sampled from a standard normal distribution $N(0,1)$, representing a random point in the latent space. Through a multi-layer neural network, the generator maps the random noise z into generated data $G(z)$ with a specific distribution and structure, aiming to produce samples that closely match the distribution of real data X_{real} so that the discriminator cannot distinguish between real and generated data.with a specific distribution and structure, aiming to produce samples that closely match the distribution of real data X_{real} so that the discriminator cannot distinguish between real and generated data[60][61]

The discriminator, on the other hand, receives both real data X_{real} and generated data $G(z)$ as inputs and outputs a probability value $D(X) \in [0,1]$, representing the confidence that the input sample is real. Specifically, $D(X)=1$ indicates that the input is real data, while $D(X)=0$ indicates that the input is generated data. The objective of the discriminator is to distinguish real data from generated data as accurately as possible.

According to the workflow illustrated in **Figure 3**, the main steps of GAN training are as follows: First, the GAN takes a random noise vector z as input, which is typically sampled from a standard normal distribution $N(0,1)$ and serves as the input to the generator. Next, the generator G processes the random noise z through a multi-layer neural network, mapping it into generated data $G(z)$ with a specific distribution and structure, aiming to resemble the real data X_{real}. Finally, the discriminator

D receives both real data X_{real} and generated data $G(z)$ and outputs a probability value representing the confidence that the input sample is real. Through this adversarial learning process, both the generator and the discriminator are continuously optimized. Eventually, the generator is capable of producing samples that closely match the real data distribution, making it increasingly difficult for the discriminator to differentiate between real and generated data.

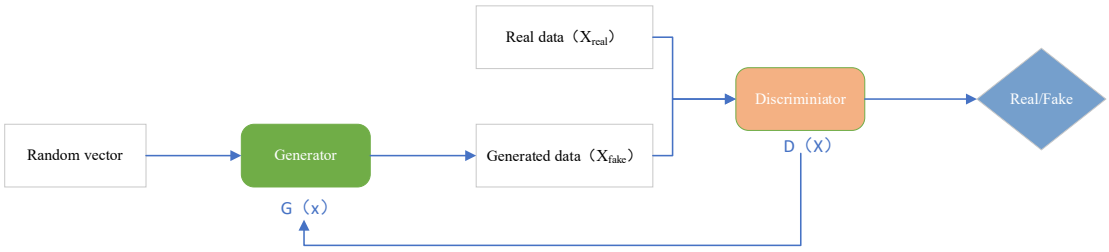

Figure 3. Workflow of the GAN Model

(2) Adversarial Learning Mechanism

During the training process, the generator and the discriminator engage in a continuous adversarial game[60][61]

1）Generator's Objective: he goal of the generator is to deceive the discriminator, making the generated data *G(z)* indistinguishable from real data. The optimization objective of the generator is to minimize the discriminator's ability to correctly classify generated samples as fake.

2）Discriminator's Objective: The discriminator strives to accurately differentiate between real and generated data. Its optimization goal is to minimize classification errors for real samples while maximizing classification errors for generated samples.

The training objectives of both networks are defined through loss functions：

$$\min_G \max_D E_{x\sim preal}[\log D(x)] + E_{z\sim pz}[\log(1-D(G(z)))] \quad (3)$$

Where The first term represents the discriminator's prediction loss on real data. The second term corresponds to the generator's attempt to deceive the discriminator.

3.1.4 VAE-GAN Model

To combine the advantages of Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), researchers proposed the VAE-GAN model. The VAE-GAN model introduces the adversarial training mechanism of GAN into the VAE framework, enabling the generator to not only produce diverse samples but also generate high-quality images. The core idea of the VAE-GAN model is to leverage the latent variable model of VAE to learn the underlying distribution of the data, while utilizing the adversarial training mechanism of GAN to enhance the quality of the generated data [62].

(1)Principles and Architecture of the Variational Autoencoder-Generative Adversarial Network (VAE-GAN)

The Variational Autoencoder-Generative Adversarial Network (VAE-GAN) seamlessly integrates the advantages of Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), aiming to explore the latent distribution of data through VAE's latent variable model while leveraging GAN's adversarial training mechanism to enhance the realism of generated data. The VAE-GAN model consists of three core components: the encoder (E), the generator (G), and the discriminator (D). Its collaborative workflow, as illustrated in **Figure 4**, can be summarized as follows [62][10]:First, the model take s defect image samples as input. The encoder E extracts the latent representation Z_E, while, to introduce sample diversity, a random vector Z_P is generated and fed into the generator G. The encoder E maps the input image to the latent space, generating the feature vector Z_E to capture the underlying structural information of the image. Meanwhile, the random noise vector Z_P is used to simulate possible defect distributions. The generator G utilizes both Z_E and Z_P to reconstruct or generate images.Finally, the generator G produces two types of images: reconstructed images and synthetic (fake) images, derived from Z_E and Z_P, respectively. The discriminator D evaluates the output of the generator by distinguishing between real and fake images while also assessing the quality of reconstructed images. The "real/fake" score produced by the discriminator serves as the adversarial loss for the GAN, ensuring that the generated images become increasingly indistinguishable from real samples.

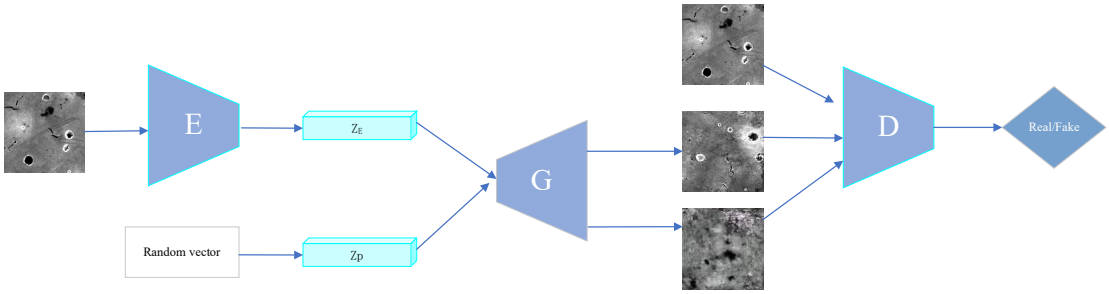

Figure 4. Workflow Diagram of VAE-GAN

(2) Loss Functions [62][10]

The training of VAE-GAN combines the reconstruction loss of VAE and the adversarial loss of GAN. The loss function consists of the following components:

- Reconstruction Loss: Ensures that the generator can accurately reconstruct the input image.
- Kullback-Leibler (KL) Divergence: Constrains the latent distribution generated by the encoder to approximate a standard normal distribution.
- Adversarial Loss: The adversarial loss is used to enhance the realism of the generated images. The generator aims to minimize the discriminator's probability of correctly classifying generated images, while the discriminator aims to maximize its probability of correctly classifying real images and minimize its probability of classifying generated images as real.
- By jointly optimizing the loss functions of the encoder, generator, and discriminator, the model can generate high-quality images while accurately capturing defect features.

By jointly optimizing the loss functions of the encoder, generator, and discriminator, the model can generate high-quality images while accurately capturing defect features.

3.1.5 Multi-VAE-GAN Algorithm

To further enhance the generative capability of the VAE-GAN model, this paper proposes an improved model that incorporates a multi-head attention mechanism. The multi-head attention mechanism has demonstrated significant advantages in generative models, particularly in capturing global dependencies within images. Mathematically, the input sequence $X \in R^{n \times d_{\text{model}}}$ (where n is the sequence length and d_{model} is the feature dimension) is first linearly projected into three matrices: Query (Q), Key (K), and Value (V) through learnable weight matrices:

$$Q = XW^{Q}, K = XW^{K}, V = XW^{V} \quad (4)$$

where $W^{Q}, W^{K}, W^{V} \in R^{d\text{model} \times dk}$ are trainable parameters, and d_k is the dimension of each head.

The projected matrices are then split into h parallel heads along the feature dimension:

$$Q_i = Split_Q(Q), K_i = Split_K(K), V_i = Split(V)(i = 1,2...h) \quad (5)$$

Each head independently computes the scaled dot-product attention:

$$Attention(Q_i, K_i, V_i) = \text{softmax}(\frac{Q_i, K_i^T}{\sqrt{d_k}})V_i \quad (6)$$

where $\sqrt{d_k}$ scales the dot products to prevent gradient vanishing.

The outputs of all heads are concatenated and linearly transformed to produce the final multi-head attention features:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, ..., head_h)W^{O} \quad (7)$$

where $W^{O} \in R^{h \cdot d_k \times d_{\text{model}}}$ is the output projection matrix.

By computing multiple attention heads in parallel, each attention head can independently focus on different regions or features of the image, enabling the model to more comprehensively understand the semantic information of the image. This mechanism not only strengthens the model's ability to model global structures but also captures fine-grained details, allowing the generative model to better maintain coherence and consistency when producing high-quality images. Furthermore, by fusing the outputs of multiple attention heads, the multi-head attention mechanism effectively integrates visual features at different levels, further enhancing the expressive power of the generative model. This parallel computation approach not only improves the model's efficiency but also makes it more flexible and robust when handling complex images, leading to outstanding performance in tasks such as image generation and image inpainting. The structure of the model is illustrated in **Figure 5**.

In the figure, Linear_Q, Linear_K, and Linear_V represent linear transformations (fully connected layers) that map the input sequence X to the query (Query), key (Key), and value (Value) spaces, respectively. The input X has a dimension of (n×dmodel), where n is the sequence length and dmodel is the model dimension. Specifically, the input sequence is transformed into Q,K, and V matrices through these three linear transformations. Subsequently, Split_Q, Split_K, and Split_V denote the division of the Q, K, and V matrices along the feature dimension into multiple heads, with each head responsible for capturing information from different subspaces of the input sequence. After division, each head independently computes attention scores and produces corresponding outputs through the scaled dot-product attention mechanism. The Concat operation then concatenates the outputs of all heads, restoring the original dimension. Finally, Linear_out is a linear transformation layer that further maps the concatenated result to generate the final output of the multi-head attention mechanism. The collaborative operation of these components enables the multi-head attention mechanism to capture diverse information from the input sequence in parallel, thereby enhancing the model's expressive power.

Although multi-head attention enables the model to attend to different subspaces in parallel and enhances its expressive power, studies have shown that using too few or too many heads does not necessarily yield optimal performance. When the number of heads is fewer than eight (e.g., 4 – 6), each head receives a higher dimensionality

(d_{model}/h), which may lead to a low-rank bottleneck, thereby limiting representational capacity and potentially degrading model performance[64]. Conversely, using more than eight heads (e.g., 12 or 16) can increase subspace redundancy and parallelism during training. However, extensive empirical evidence reveals that many of these heads are redundant at inference time, with only a few contributing significantly to the final output. These redundant heads can often be pruned without noticeable loss in performance.

For instance, in the WMT English – Russian translation task, pruning 38 out of 48 heads resulted in only a 0.15 drop in BLEU score[65]. Michel et al. also observed that even after removing most of the 16 heads, the model maintained near-original performance, stating: "Even when most heads are removed at test time, BLEU remains nearly constant," and in some layers, a single head was sufficient. This suggests that configurations with more than eight heads often introduce unnecessary parameters and computational overhead without improving representational diversity or generalization.

Moreover, recent studies have proposed mechanisms such as Grouped Head Attention and Voting-to-Stay, which aim to identify and prune redundant heads, further confirming the diminishing returns and prune-ability of excessive heads[66]. Therefore, using eight heads is often considered a "sweet spot": it provides sufficient subspace diversity while avoiding dimensional bottlenecks and excessive redundancy, making it an empirically optimal trade-off between performance and efficiency. This configuration is widely adopted in vision Transformer models like ViT[67], where it consistently yields strong results.

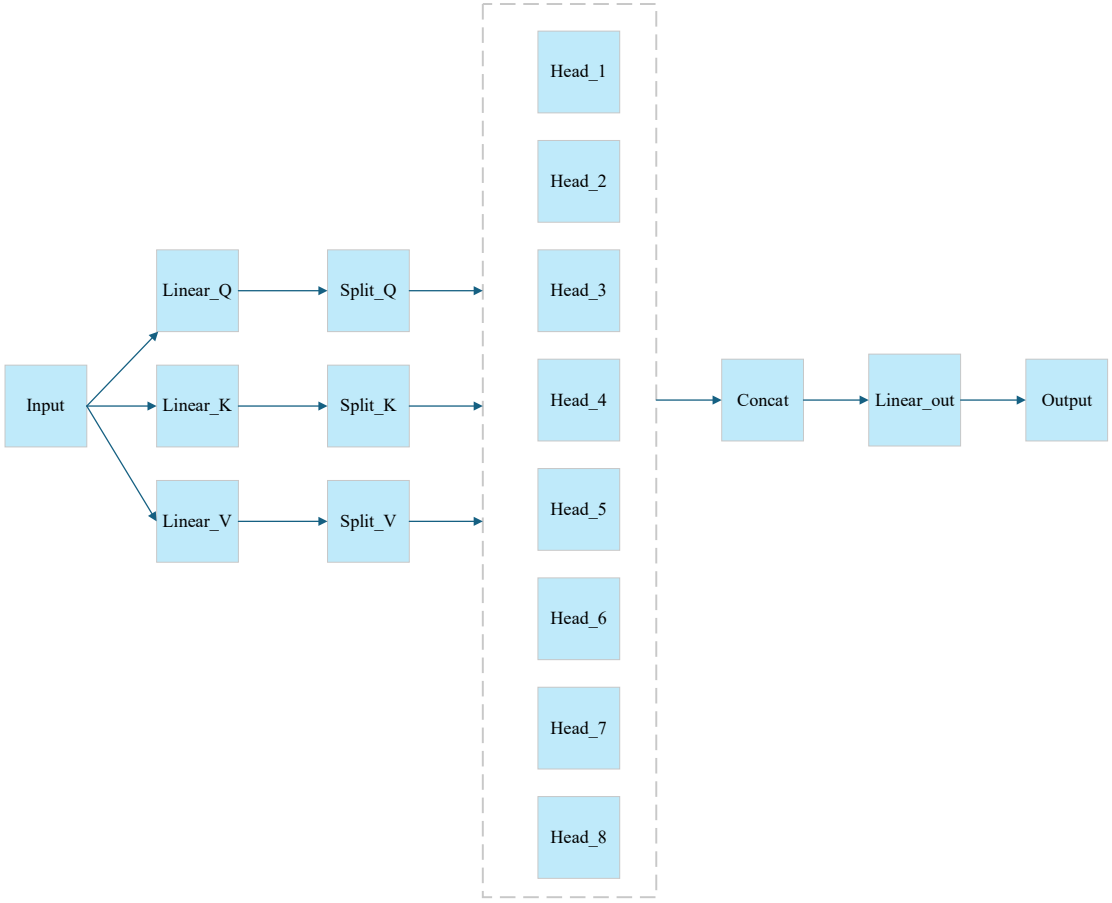

Figure 5. Structure Diagram of the Multi-Head Attention Mechanism

The model proposed in this paper improves upon the VAE-GAN by incorporating the multi-head attention mechanism into the encoder and generator, significantly enhancing the model's performance. As shown in **Figure 6**, the overall architecture of the model retains the advantages of the original VAE-GAN while achieving more effective feature extraction and image generation through the introduction of the multi-head attention mechanism.

In the encoder section, the traditional VAE structure primarily relies on convolutional operations to extract local features. In contrast, the multi-head attention mechanism introduced in this paper captures long-range dependencies between different regions of the input image, thereby better extracting global features. By computing multiple attention heads in parallel, the model can simultaneously focus on

different feature subspaces of the image, not only enhancing the comprehensiveness of feature extraction but also improving the model's ability to understand complex image structures.

In the generator section, the introduction of the multi-head attention mechanism also brings significant performance improvements. Traditional GAN generators mainly rely on transposed convolutional operations to progressively generate images, which can easily result in a lack of global consistency in the generated images. By integrating the multi-head attention mechanism into the generator, the proposed model dynamically adjusts the relationships between different regions during the generation process, thereby producing more realistic and diverse images. The multi-head attention mechanism enables the generator to better capture global structural information in the image, avoiding the common mode collapse issue in traditional methods while improving the detail quality and visual realism of the generated images.

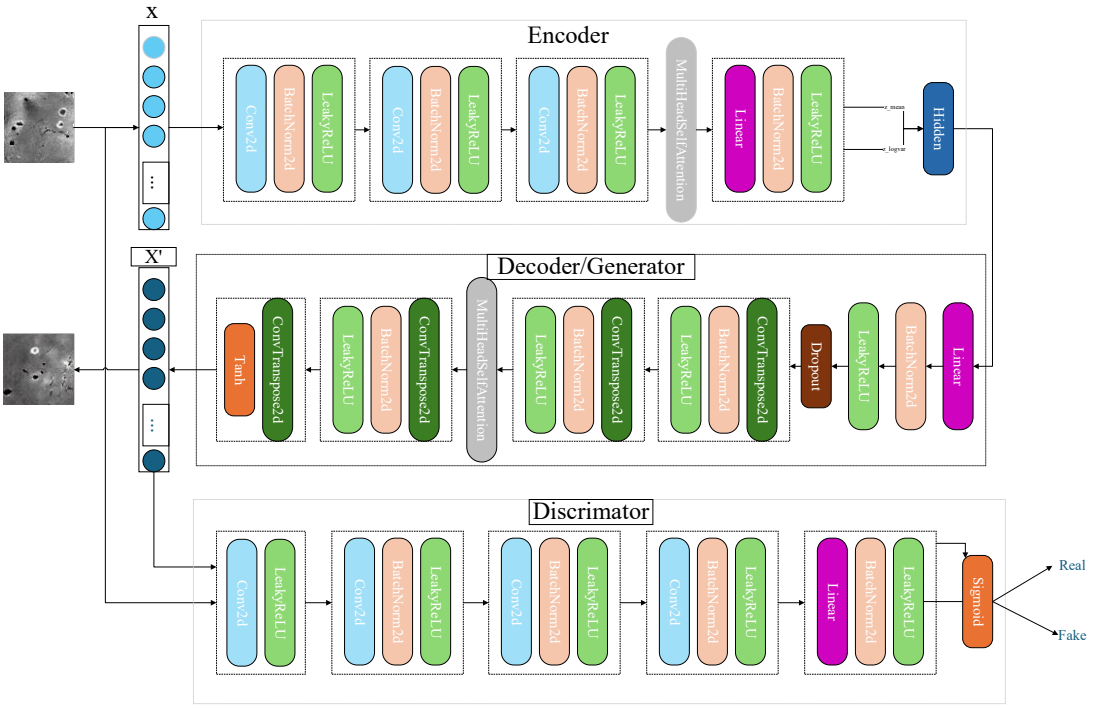

Figure 6. Flowchart of the proposed Multi-VAE-GAN model

3.2. Defect Detection Models

In this study, several mainstream object detection models were selected for training and testing to comprehensively evaluate their performance in defect detection tasks. The specific models chosen include the following:

(1) SSD

SSD is an efficient single-stage object detection model that generates multi-scale candidate bounding boxes directly on feature maps and performs classification simultaneously. It offers high detection speed, making it particularly suitable for real-time defect detection applications.

(2) YOLO Series Models

YOLOv4: YOLOv4 introduces significant improvements in network architecture optimization, data augmentation strategies, and training process tuning. It achieves a good balance between detection accuracy and speed, making it suitable for defect detection scenarios that require both real-time performance and accuracy.

YOLOv7: YOLOv7 further optimizes model efficiency and detection accuracy, particularly excelling in complex scenes and small object detection. It is well-suited for defect detection tasks with high demands for both real-time performance and precision.

YOLOv8: YOLOv8 introduces a redefined architecture that prioritizes scalability and efficiency. It adopts an anchor-free detection mechanism, streamlined feature fusion pathways, and adaptive training strategies such as dynamic label assignment. These innovations enhance both precision and inference speed across diverse model scales (from nano to extra-large), making it particularly effective for defect detection tasks—even in challenging scenarios involving small or occluded defects.

(3) Faster R-CNN

Faster R-CNN is a two-stage object detection model based on region proposals. It generates high-quality candidate bounding boxes through a Region Proposal Network (RPN) and combines them with a fully convolutional network (CNN) for precise classification and regression. Its core advantage lies in its high detection

accuracy, making it suitable for scenarios where detection results are critical.

During the model training and testing phases, this study conducts a comprehensive evaluation of the aforementioned models' performance in defect detection tasks. The evaluation metrics include Precision, Recall, and mean Average Precision to meet the requirements of practical applications.

4. Experiment and Results Analysis

4.1. Experimental Environment and Parameter Configuration

The experiments were conducted on the AutoDL cloud computing platform. The hardware configuration of the cloud environment includes a 64-bit Windows 10 operating system, an Intel® Xeon® Platinum 8481C CPU @ 2.00 GHz, 16GB of RAM, and an NVIDIA RTX 4090D GPU with 24GB of dedicated memory. The deep learning software environment is set up with Python 3.8, utilizing CUDA 11.8 for GPU acceleration, and implemented using the PyTorch 2.0.0 framework.

In this section, we apply the algorithmic steps outlined in Section 3.1 to the defect dataset to evaluate the effectiveness of the proposed method. Prior to conducting the experiments, we optimized the VAE-GAN model by configuring the batch size to 16, the total number of epochs to 800, and the learning rate to 3e-4, using the Adam optimizer, as detailed in Table 1.

Table 1. Model Parameter Settings

Parameter settings	Details
Batch_size = 16	Number of batch processes is 16
Epoch = 800	Training 800 rounds of data
Learing_rate = 3e-4	Initial learning rate of 3e-4
Adam	Optimizer

4.2. Data Collection

To train and validate the model, scanning electron microscope (SEM) images of additively manufactured components were utilized. A total of 11 laser powder bed fusion (L-PBF) Ni939 sample SEM images, each with a resolution of 1024×960 pixels, were collected from the SEM image archive. Each of these images contains both porosity and crack defects. The dataset was randomly divided into a training set and a test set at a ratio of 7:3.

Initially, defects were annotated, and based on their annotated locations, the images were cropped into multiple 128×128 pixel patches to construct the raw dataset. In the training set, the number of porosity and crack samples was 1,551 and 545, respectively. All experiments were conducted on the training set, while the test set remained unchanged. However, the dataset exhibits a relatively small sample size and an inherent class imbalance, as the number of the two defect types differs significantly. In subsequent steps, various data augmentation techniques will be employed to mitigate these issues.

4.3. Evaluation Metrics

In this study, a comprehensive evaluation was conducted to assess both the generative performance of the proposed Mult-VAE-GAN model and the accuracy of the defect detection model. The quality of generated defect samples was measured using commonly adopted objective metrics in the image generation domain, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), along with subjective assessment methods focusing on defect-specific characteristics. Meanwhile, the accuracy of the defect detection model was evaluated using classification performance metrics, including accuracy, precision, and recall. These metrics were used to analyze the effectiveness of the generated samples in downstream defect detection tasks and to assess the practical applicability of the model.

The selection of these evaluation metrics aims to provide a scientific and comprehensive assessment of image quality, realism, and the impact of generated images on real-world defect detection applications.

4.3.1 Data Augmentation Evaluation Metrics

(1) The Peak Signal-to-Noise Ratio (PSNR) is one of the conventional metrics for assessing imagequality. A higher PSNR value indicates smaller differences between the generated image and the reference image, implying better quality of the generated image. The definition of PSNR is as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right) \tag{8}$$

Where L represents the maximum pixel value of the image (for an 8-bit image, L=255).Mean Squared Error (MSE) measures the average squared difference between the generated image and the reference image. It is calculated as follows:

$$MSE = \frac{1}{H \cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W}(I(i,j) - \hat{I}(i,j))^2 \tag{9}$$

Where N=H×W, where H and W represent the height and width of the image, respectively. $I(i.j), \hat{I}(i,j)$ denote the pixel values at position (i, j) · in the reference image and the generated image, respectively.

(2)The Structural Similarity Index Measure (SSIM) is a metric used to evaluate the similarity between two images in terms of luminance, contrast, and structure. It is designed to more accurately reflect image quality as perceived by the

human visual system. The formula for calculating SSIM is as follows:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (10)$$

Where μ_x and μ_y represent the mean values of images x and y, respectively; σ_x^2 and σ_y^2 denote the variances of images x and y, respectively; σ_{xy} is the covariance between images x and y;$C_1=(K_1L)^2$ and $C_2=(K_2L)^2$ are stabilization constants, where L is the maximum pixel value, K1 and K2 are small positive numbers used to prevent division by zero.

4.3.2 Evaluation Metrics for Defect Detection

To accurately evaluate the performance of the defect detection model, this paper adopts Average Precision (AP) and mean Average Precision (mAP) as evaluation metrics. AP is a comprehensive metric calculated based on Precision and Recall, while mAP is the average AP value across all target categories. These two metrics are widely used in the field of object detection and can comprehensively reflect the model's performance in defect detection tasks.

(1) Precision and Recall

Precision and Recall are the foundational concepts for defining AP and mAP. Their formulas are as follows:

Precision is defined as:

$$precision = \frac{TP}{TP + FP} \quad (11)$$

Recall is defined as:

$$recall = \frac{TP}{TP + FN} \quad (12)$$

Where TP (True Positives): The number of samples correctly predicted as positive.FP (False Positives): The number of samples incorrectly predicted as positive.FN (False Negatives): The number of samples incorrectly predicted as negative.

(2) Average Precision (AP)

AP is obtained by calculating the area under the Precision-Recall (PR) curve. Its formula is as follows:

$$AP = \int_0^1 \Pr ecision(\mathrm{Re}\, call)d\,\mathrm{Re}\, call \quad (13)$$

In practical calculations, a discretization method is typically employed. The precision and recall values are progressively accumulated based on the confidence scores of the detection boxes, and the integral value of the curve is estimated using interpolation. AP can evaluate the detection performance of the model on a single defect category.

(3) Mean Average Precision (mAP)

mAP is a comprehensive metric obtained by averaging the AP values across all target categories. The formula is as follows:

$$mAP = \frac{1}{K}\sum_{i=1}^{k} AP_i \quad (14)$$

Where N: The number of target categories. AP_i: The average precision of the i-th category.

mAP serves as a comprehensive evaluation criterion for multi-category detection performance. A higher mAP value indicates stronger overall detection capability of the model.

4.4 Experimental Results and Discussion

In this section, extensive experiments are conducted to demonstrate the effectiveness of the proposed Mult-VAE-GAN algorithm in augmenting additive manufacturing defect sample data and improving defect detection accuracy. The key experiments include analyzing the impact of data preprocessing, evaluating the performance of Mult-VAE-GAN, and validating the effectiveness of the proposed method based on SSIM and PSNR. Additionally, multiple defect detection models are utilized to verify that the data augmentation approach enhances defect detection accuracy.

4.4.1 Data Preprocessing

In the original dataset, there exists a significant class imbalance among defect samples, which adversely affects the performance of defect detection. Specifically, when using the YOLOv4 algorithm for initial defect detection, the average precision (AP) for cracks was 39.20%, while for pores, it reached 90.51%, resulting in an overall mAP of 64.85%. This outcome suggests that the imbalance in data distribution is a primary factor contributing to the lower detection performance for cracks.

To mitigate the performance degradation caused by class imbalance, conventional data augmentation techniques were applied to expand and balance the dataset. The augmentation process included 90° and 180° rotations, as well as adjustments to brightness, contrast, saturation, and hue. These transformations preserved the original defect characteristics while increasing data diversity. After augmentation, the number of pore samples increased to 1,656, and the number of crack samples increased to 1,388, thereby alleviating the imbalance issue to some extent.

This data preprocessing step significantly improved the balance of the dataset, providing a more reliable foundation for subsequent data augmentation and defect detection model training.

4.4.2 Data Augmentation

During the defect recognition experiments, it was observed that the imbalance in defect sample distribution and the limited overall dataset size negatively impacted the performance and accuracy of the detection model. To

address this issue, an initial dataset balancing process was conducted to ensure a more even distribution of defect categories. Additionally, traditional data augmentation techniques, such as rotation, flipping, and scaling, were applied. However, when defect detection was performed using only YOLOv4 after traditional data augmentation, the detection performance did not improve significantly; instead, a downward trend was observed. Specifically, the detection accuracy of YOLOv4 decreased for both Pore and Crack defect categories, resulting in an overall mAP of only 59.45%. This decline in performance may be attributed to the fact that traditional data augmentation methods potentially damaged defect features, leading to degraded detection results and affecting the model's learning or generalization ability. Moreover, traditional data augmentation methods may not have effectively enhanced the diversity of defect features, preventing the model from sufficiently learning the key characteristics of different defect types.

To overcome these limitations, we employed the improved Mult-VAE-GAN data augmentation model to generate additional high-quality reconstructed images. This method is designed to enhance the generalization ability of the detection model and improve overall detection performance in additive manufacturing.

To assess the quality of the generated reconstructed images, this study adopted two widely used evaluation metrics: SSIM and PSNR. The generated images were evaluated to ensure their quality before being incorporated into the balanced dataset. Ultimately, high-quality reconstructed images were selected and integrated into the dataset to further enrich the data and enhance classification accuracy.

Table 2 presents the distribution of defect samples across the original dataset, the dataset after balancing, and the dataset after augmentation using Mult-VAE-GAN.

Table 2. Defect Sample Distribution

	Pore	Crack
Raw Data	1551	545
Data Balancing	1656	1388
Mult-VAE-GAN Data Augmentation	3844	3280

Figure 7 presents the experimental results obtained using two image quality evaluation metrics: Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The left subplot illustrates the PSNR evaluation results, while the right subplot displays the SSIM evaluation results.

From **Figure 7**, it can be observed that as the data augmentation process progresses, the values of both PSNR and SSIM exhibit a steady upward trend, indicating a continuous improvement in the quality of the reconstructed images. Notably, during the initial phase of data augmentation, the PSNR and SSIM values increase rapidly, suggesting that the Mult-VAE-GAN model can quickly capture the primary structural and textural information of the images in the early training stage, thereby generating images of relatively high quality. This rapid increase may be attributed to the initialization of model parameters and the early adjustments made by the optimization algorithm, allowing the generated images to achieve satisfactory visual quality within a limited number of iterations.

To quantitatively validate the superiority of the proposed Mult-VAE-GAN, we compare its performance with the original VAE-GAN framework. As shown in Table 3, our model achieves significant improvements, with SSIM and PSNR values increasing from 0.7365 to 0.8214 (+11.5%) and from 21.4694 to 22.5437 (+5.0%), respectively. These results confirm that the multimodal architecture enhances both structural coherence and pixel-level fidelity compared to the baseline.

Table 3. Defect Sample Distribution

	SSIM	PSNR
VAE-GAN	0.7365	21.4694
Mult-VAE-GAN	0.8214	22.5437

As the data generation process continues, the growth rate of PSNR and SSIM gradually slows down and stabilizes in the later stages of training. The PSNR value ultimately stabilizes around 22, while the SSIM value remains approximately 0.8, indicating that the generated images not only maintain a high signal-to-noise ratio at the pixel level but also effectively preserve the semantic information of the original images in terms of structural perception. Furthermore, although minor fluctuations in the numerical values occur in the later training stages, the overall trend remains stable at a relatively high level. This suggests that the model is capable of maintaining image quality over extended training periods and demonstrates a certain degree of generalization, enabling it to adapt to variations in data distribution.

To further enhance the quality and balance of the dataset, we performed quality evaluation-based filtering on the reconstructed images and incorporated the selected high-quality samples into the original dataset. This optimization strategy not only increases the diversity of the dataset but also provides high-quality input data for subsequent defect detection model training, thereby effectively improving classification accuracy and generalization performance.

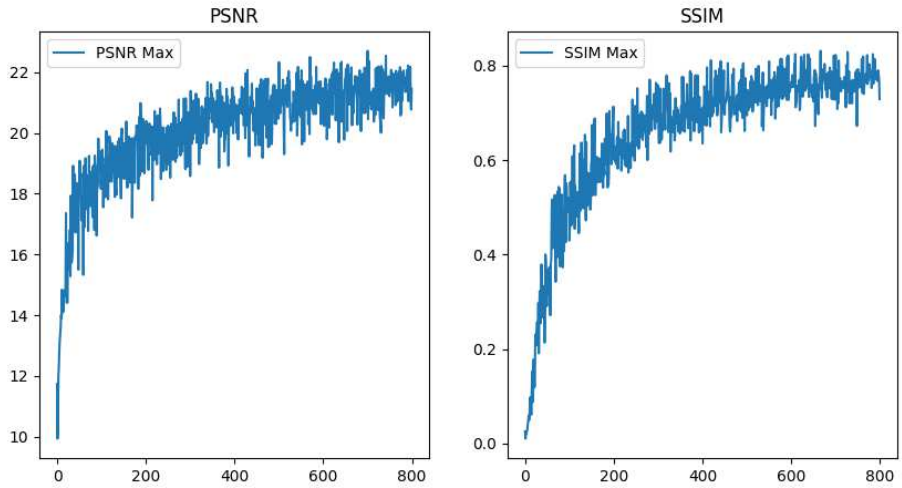

Figure 7. Evaluation of Reconstructed Images

4.4.3 Defect Detection

To enhance the generalization capability and detection accuracy of defect detection models, this study proposes a Mult-VAE-GAN-based data augmentation method to generate a more diverse and complex set of defect samples. Using the augmented dataset, we retrained and validated five mainstream object detection models: YOLOv7, YOLOv4, YOLOv8, SSD, and Faster R-CNN.

Before introducing the Multivariate Variational Autoencoder Generative Adversarial Network (Mult-VAE-GAN) for data augmentation, we conducted defect detection experiments on two defect types—pore and crack—using a balanced dataset without any augmentation. Table 4 presents the detection performance of various methods at this stage.

According to the results, SSD achieved the best performance in crack detection, with an average precision (AP) of 58.19%, and also showed a competitive AP of 89.78% for pore detection. This resulted in the highest overall mean average precision (mAP) of 73.99% among all methods. Faster R-CNN yielded the highest AP for pore detection (92.07%), but its performance on crack detection was slightly lower (46.85%), leading to a mAP of 69.46%. YOLOv7 demonstrated balanced performance across both defect types, with an AP of 47.26% for crack and 90.86% for pore, resulting in a mAP of 69.06%. YOLOv8 showed comparable results, with a mAP of 67.07%. In contrast, YOLOv4 exhibited the weakest performance on crack detection (AP of 39.02%), and although it achieved 91.43% AP for pore detection, its overall mAP was the lowest at 65.18%.

In summary, these baseline detection results provide a reference for evaluating the effectiveness of the proposed Mult-VAE-GAN data augmentation technique in enhancing defect detection performance in subsequent experiments.

Table 4. Defect Detection Results Before Data Augmentation.

Method	AP（%）		mAP（%）
	Pore	Crack	
Yolov4	91.43%	39.02%	65.18%
Yolov7	90.86%	47.26%	69.06%
Yolov8	90.46%	43.68%	67.07%
SSD	89.78%	58.19%	73.99%
Faster-RCNN	92.07%	46.85%	69.46%

After incorporating the Mult-VAE-GAN data augmentation technique, we retrained and tested the same balanced dataset. The results, presented in Table 5, indicate that data augmentation significantly improved the performance of all detection methods compared to the pre-augmentation results shown in Table 4.

Table 5 reports the average precision (AP) for Pore and Crack detection, as well as the mean average precision (mAP) for each method after data augmentation. The key findings are as follows:

- Faster R-CNN exhibited the most prominent improvement in overall performance. Its average

precision (AP) for Crack detection surged from 46.85% to 84.67%, while Pore detection experienced a slight decline from 92.07% to 90.82%. Nonetheless, the mean average precision (mAP) increased significantly from 69.46% to 87.75%, representing an improvement of 18.29 percentage points, indicating a remarkable enhancement in detection capability.

- SSD also demonstrated a dramatic performance boost, particularly in Crack detection, where the AP rose from a previously lowest 50.90% to 83.47%. The AP for Pore detection remained relatively stable, slightly decreasing from 89.78% to 89.18%. However, the overall mAP increased from 73.99% to 86.33%, marking the largest relative gain among all models (+12.34%).
- YOLOv4 achieved a significant improvement in Crack detection AP, which rose from 39.02% to 80.69%, nearly doubling its previous performance. Although its Pore AP slightly declined from 91.43% to 89.74%, the overall mAP experienced a sharp increase from 65.18% to 85.21%, an improvement of 20.03 percentage points, the highest among all models.
- YOLOv8 showed consistent performance gains across both defect categories. The AP for Crack detection increased from 43.68% to 80.72%, while Pore detection saw a moderate decline from 90.46% to 88.60%. Despite this, the overall mAP improved from 67.07% to 84.66%, resulting in a 17.69% enhancement.
- YOLOv7 maintained robust performance following augmentation. The AP for Pore increased from 90.86% to 92.22%, whereas the Crack AP slightly decreased from 47.26% to 76.13%. Even so, the model's mAP improved from 69.06% to 84.18%, indicating a 15.12% performance gain.

In summary, the comparative results clearly demonstrate that the Mult-VAE-GAN data augmentation technique effectively enhances the performance of defect detection models. The ability of Mult-VAE-GAN to generate structurally diverse and representative Crack samples helps mitigate the challenges posed by insufficient data and structural variability, thereby significantly improving the learning capacity and generalization of detection models.

From the perspective of pore defect detection, the improvements observed are not uniformly significant. One possible explanation is that the baseline model had already achieved a relatively high detection accuracy (typically around 90%) prior to data augmentation, potentially reaching a performance saturation point. Consequently, the marginal benefit from further augmentation becomes limited. As noted by Maayan Frid-Adar et al.[63], increasing the number of training examples generally improves performance; however, once saturation is reached, adding more augmented samples may no longer yield noticeable gains in classification results. When the classification performance is already high, the additional benefits from augmentation tend to be minimal. This phenomenon does not necessarily indicate the failure of the augmentation method itself, but rather reflects the presence of strong intrinsic features in the original pore images, which diminish the incremental value of augmented data. Furthermore, the multi-VAE-GAN approach, while effective in many cases, may introduce subtle artifacts such as edge blurring or texture distortion when handling fine-grained defects. These minor variations can present adaptation challenges for certain models, leading to slight fluctuations in recognition accuracy.

However, considering the comprehensive metric of mean Average Precision (mAP), all models exhibit consistent improvement, further validating the effectiveness and generalization capability of the Multi-VAE-GAN augmentation strategy at a macro level.

Table 5. Defect Detection Results After Data Augmentation.

Method	AP（%）		mAP（%）
	Pore	Crack	
Yolov4	89.74%	80.69%	85.21%
Yolov7	92.22%	76.13%	84.18%
Yolov8	88.60%	80.72%	84.66%
SSD	89.18%	83.47%	86.33%
Faster-RCNN	90.82%	84.67%	87.75%

By comparing Table 4 and Table 5, it is evident that data augmentation effectively enhances the accuracy of defect detection models while also improving the robustness of different detection methods. This provides strong technical support for high-precision defect detection in real-world applications.

To further evaluate the impact of the Mult-VAE-GAN data augmentation technique on defect detection performance, we conducted predictions on selected test samples and generated corresponding detection result images. These visualizations provide a clear depiction of each model's ability to detect Pore and Crack defects.

(1) YOLOv7 Detection Results

YOLOv7 exhibited the superior performance after data augmentation. As shown in **Figure 8**, the model accurately locates and identifies both Pore and Crack defects with minimal false positives or false negatives. Notably, YOLOv7 demonstrated strong robustness in

detecting cracks within complex backgrounds, maintaining high precision even in challenging scenarios.

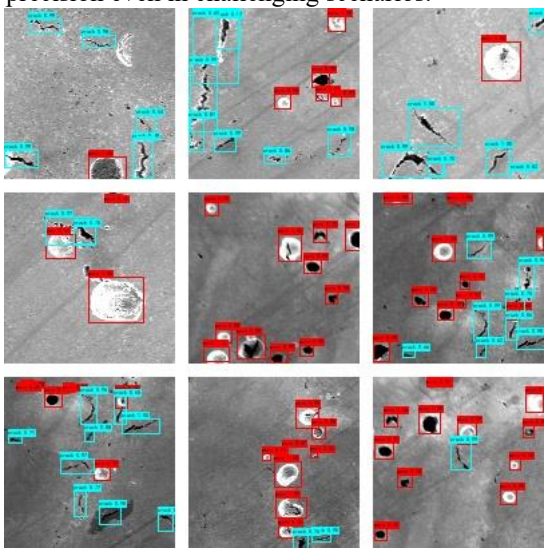

Figure 8. YOLOv7 Detection Results

(2) YOLOv4 Detection Results

YOLOv4 showed noticeable improvement following data augmentation; however, its detection results indicate that some minor cracks were still missed. As illustrated in **Figure 9**, while YOLOv4 effectively detects Pore defects, its performance in detecting fine cracks is slightly inferior to YOLOv7. Nevertheless, compared to the pre-augmentation results, YOLOv4's overall detection accuracy has improved significantly.

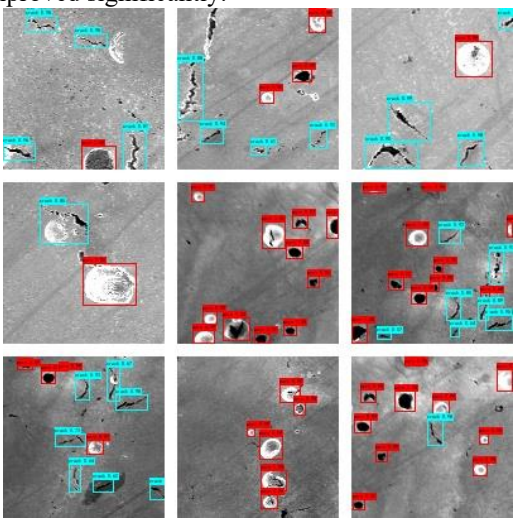

Figure 9. YOLOv4 Detection Results

(3) SSD Detection Results

Among all models, SSD exhibited the most substantial improvement after data augmentation, particularly in Crack detection. As depicted in **Figure** 10, SSD effectively identifies Crack defects, and its false detection rate is significantly reduced. However, compared to YOLOv7 and YOLOv4, SSD still experiences missed detections in Pore detection, especially when the Pores are small or densely distributed.

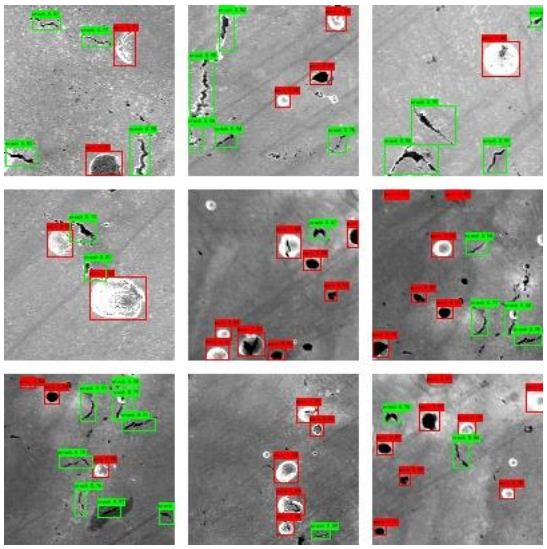

Figure 10. SSD Detection Results.

(4) Faster R-CNN Detection Results

Faster R-CNN also showed performance enhancements following data augmentation. As presented in **Figure 11**, the model accurately detects Pore and Crack defects. However, in complex backgrounds, a few false detections still occur, particularly in Crack detection.

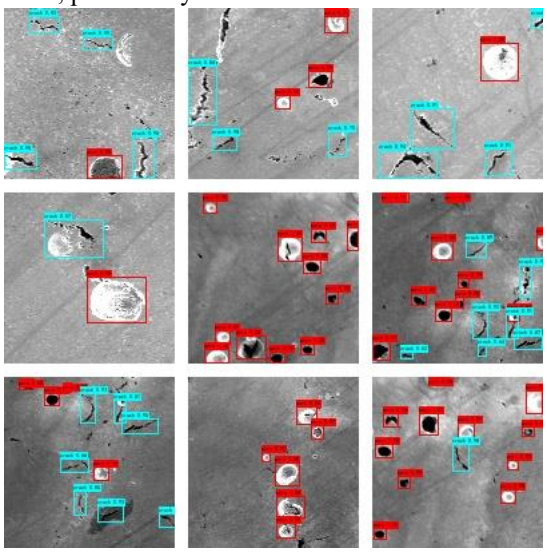

Figure 11. Faster R-CNN Detection Results

(5) Yolov8 Detection Results

The implementation of data augmentation has led to an enhancement in the overall performance of YOLOv8. As illustrated in **Figure 12**, the model demonstrates high accuracy in detecting hole and crack defects. Nevertheless, in complex backgrounds, minor false positives may still occur, particularly for small pores and fine cracks.

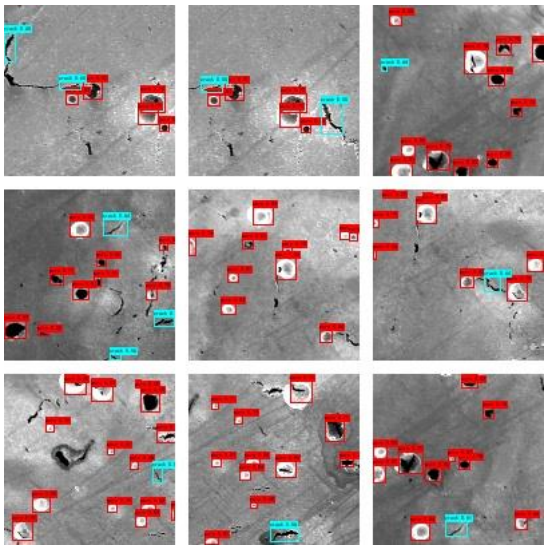

Figure 12. Yolov8 Detection Results

Overall, the Mult-VAE-GAN data augmentation technique significantly improves defect detection performance,. These findings further confirm the effectiveness of this augmentation approach in improving defect detection accuracy and robustness across different models.

4.5 Performance Evaluation of Mult-VAE-GAN Using the NEU-DET Public Dataset

The NEU-DET dataset, developed by the research team led by Kechen Song at Northeastern University, is specifically designed for the detection and recognition of steel surface defects. This dataset comprises 1,800 images covering six common types of steel surface defects: Crazing, Inclusion, Patches, Pitted Surface, Rolled-in Scale, and Scratches, as illustrated in Figure 13.

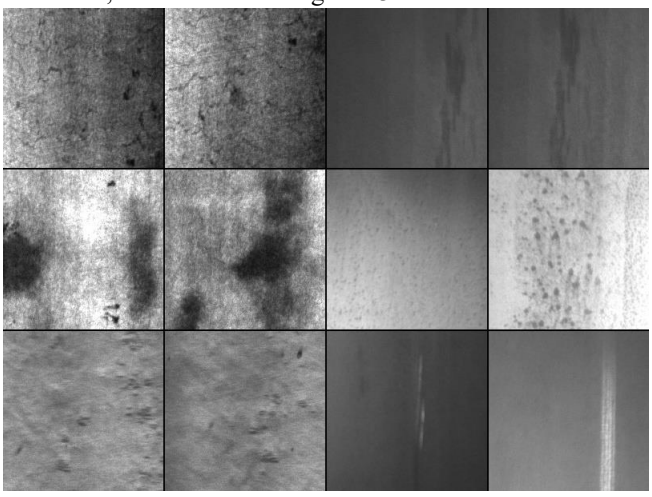

Figure 13. NEU-DET datasets

The validation experiments on the NEU-DET dataset adopted essentially the same parameters as those for the AM dataset. Nevertheless, given NEU-DET's larger data volume, the training converged stably at 400 epochs without requiring the 800 epochs needed for AM. See Table 5 for parameter configurations.

Table 5. NEU-DET dataset experimental parameters

Parameter settings	Details
Batch_size = 16	Number of batch processes is 16
Epoch = 400	Training 400 rounds of data
Learing_rate = 3e-4	Initial learning rate of 3e-4
Adam	Optimizer

Based on the experimental records and Figure 14, the SSIM values remained stable within the range of 0.76 – 0.77, peaking at 0.79, while the PSNR consistently ranged between 27 – 28, reaching a maximum of 30. Compared with the experimental results on the AM dataset, although the SSIM values showed a slight decrease, the PSNR exhibited a significant improvement of 5 points. These findings demonstrate that the Mult-VAE-GAN model is not only effective for the AM dataset but also demonstrates strong generalization capability across other domain-specific datasets.

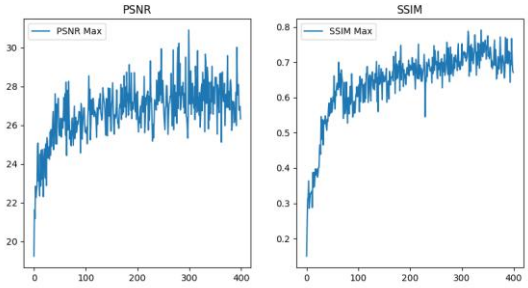

Figure 15. NEU-DET Generated Image Evaluation

Figure 15 demonstrates the synthetic images produced by the proposed Mult-VAE-GAN framework, exhibiting realistic defect patterns consistent with the training distribution.

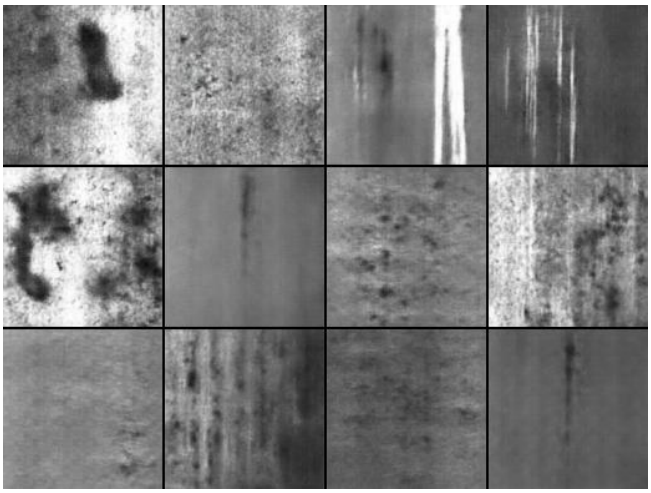

Figure 14. NEU-DET Generated Image

5. Conclusion

This study proposes a data augmentation method based on Mult-VAE-GAN, aimed at enhancing the generalization ability and detection accuracy of defect detection models in the context of additive manufacturing (AM). Due to the complex physical and chemical interactions in the AM process, defect patterns exhibit high variability and randomness. Traditional defect datasets are often limited in scale and suffer from imbalanced sample distributions, negatively impacting model performance. To address this challenge, this study utilizes Mult-VAE-GAN to generate defect samples with greater diversity and complexity, thereby improving the performance of detection models across different defect types.

Notably, although this study primarily focuses on the field of additive manufacturing, Mult-VAE-GAN, as a general data augmentation method, is equally applicable to other defect detection scenarios, such as semiconductor manufacturing, welding quality inspection, and defect detection in aerospace composite materials. Its key advantage lies in its ability to effectively expand datasets and improve model generalization across various defect patterns.

In the experimental evaluation, five state-of-the-art object detection models—YOLOv4, YOLOv7, YOLOv8, SSD, and Faster R-CNN—were employed to assess the effectiveness of the proposed Multi-VAE-GAN data augmentation technique on the detection of Pore and Crack defects. The results demonstrate that the augmented data significantly enhances detection performance across all models, particularly in terms of average precision (AP) and mean average precision (mAP). Specifically, YOLOv4 achieved a mAP improvement from 65.18% to 85.21%, with Crack detection AP increasing from 39.02% to 80.69%. YOLOv7 and YOLOv8 recorded mAP gains of 15.12 and 17.69 percentage points, respectively, with YOLOv8 showing a remarkable 37.04 percentage-point increase in Crack detection. SSD exhibited the most substantial relative improvement, with mAP rising from 73.99% to 86.33%. Faster R-CNN achieved the highest post-augmentation mAP of 87.75%, up from 69.46%. These results indicate that the Multi-VAE-GAN effectively addresses challenges associated with limited and imbalanced datasets by generating diverse defect representations, thereby improving model generalization, reducing false detections, and enhancing overall detection robustness.

This study contributes both theoretically and practically. From a theoretical perspective, it integrates the VAE-GAN generative model with defect detection tasks, providing a novel data-driven solution for defect detection in additive manufacturing. From a practical perspective, the experimental results confirm that this method effectively enhances the performance of mainstream object detection models, particularly in cases with insufficient or imbalanced datasets, significantly improving their generalization ability.

Despite the remarkable improvements achieved by the Mult-VAE-GAN data augmentation method in defect detection, several challenges remain. First, the training process requires substantial computational resources, affecting its efficiency in industrial applications. Second, the quality and diversity of generated samples still have room for further optimization. Future research directions include: (1) optimizing the training process of Mult-VAE-GAN to enhance the quality and diversity of generated samples; (2) integrating transfer learning or few-shot learning techniques to reduce computational costs and improve data utilization efficiency; and (3) exploring real-time defect detection and augmentation by incorporating Mult-VAE-GAN into online defect detection systems, enabling real-time defect data generation to enhance detection efficiency in industrial production processes.

In conclusion, the proposed Mult-VAE-GAN data augmentation method provides an effective data generation strategy for defect detection in additive manufacturing and has been experimentally validated to improve detection accuracy and generalization ability. Future research will focus on enhancing computational efficiency, optimizing the quality of generated samples, and exploring multimodal data fusion to further advance defect detection technology in additive manufacturing.

Data availability statement

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No.72471172) and State Administration of Science,Technology and Industry for National Defense (Grant No. JSZL2022204B005).

Conflict of interestd

The authors declare that there is no conflict of interests regarding the publication of this article.

References

[1] Li Jingchang,Zhou Qi,Cao Longchao,Wang Yanzhi & Hu Jiexiang.(2022).A convolutional neural network-based multi-sensor fusion approach for in-situ quality

monitoring of selective laser melting.Journal of Manufacturing Systems429-442.
[2] Wen Hao,Huang Chang & Guo Shengmin.(2021).The Application of Convolutional Neural Networks (CNNs) to Recognize Defects in 3D-Printed Parts.Materials(10),2575-2575.
[3] Z.M. Guo, P.J. Ni, Y.L. Dai, W.G. Zhang, K.D. Huang, J. Phys.: Conf. Ser. 1827(1), 012039 (2021)
[4] V.K. Nadimpalli, G.M. Karthik, G.D. Janakiram, P.B. Nagy, Int. J. Adv. Manuf. Technol. 108(5), 1793–1810 (2020)
[5] Wang Jianyong,Mu Chunyang,Mu Song,Zhu Rui & Yu Hua.(2023).Welding seam detection and location: Deep learning network-based approach.International Journal of Pressure Vessels and Piping,202,
[6] S. Niu, B. Li, X. Wang, et al., Region-and Strength-Controllable GAN for Defect Generation and Segmentation in Industrial Images, IEEE Trans. Ind. Informatics 18 (7) (2021) 4531–4541.
[7] Chung Jihoon,Shen Bo & Kong Zhenyu James.(2023).Anomaly detection in additive manufacturing processes using supervised classification with imbalanced sensor data based on generative adversarial network.Journal of Intelligent Manufacturing(5),2387-2406.
[8] Dosovitskiy, A., & Brox, T. (2016). Generating images with perceptual similarity metrics based on deep networks. arXiv preprint arXiv:1602.02644. https://arxiv.org/abs/1602.02644
[9] Mescheder, L., Nowozin, S., & Geiger, A. (2017). Adversarial variational Bayes: Unifying variational autoencoders and generative adversarial networks. arXiv preprint arXiv:1701.04722. https://arxiv.org/abs/1701.04722
[10] Y.T. Wang, Y.X. Wei, H. Wang A class imbalanced wafer defect classification framework based on variational autoencoder generative adversarial network Measurement Science and Technology, 34 (2) (2023), Article 024008
[11] T. Zhang, J. Chen, F. Li, et al., Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, ISA Trans. 119 (2022) 152–171.
[12] Waheed, A.; Goyal, M.; Gupta, D.; Khanna, A.; Turjman, F.A.; Pinheiro, P.R. CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection. IEEE Access 2020, 8, 91916.
[13] T. Zhang, S. He, J. Chen, et al., Towards small sample challenge in intelligent fault diagnosis: attention weighted multi-depth feature fusion net with signals augmentation, IEEE Trans. Instrum. Meas. 71 (2021) 3134999.
[14] deep learning-based medical image segmentation. OpenReview.net. 2018.122. Fabio P, Christina V, Sandra A, Eduardo V. Data augmentation for skin lesion analysis. In: ISIC skin image analysis workshop and challenge @ MICCAI 2018. 2018.
[15] Karzuhisa M, Akira H, Akane M, Hiroshi K. Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. In: International skin imaging collaboration (ISIC) 2017 challenge at the international symposium on biomedical imaging (ISBI). 2017.
[16] Wenyuan Cui,Yunlu Zhang,Xinchang Zhang,Lan Li & Frank Liou.(2020).Metal Additive Manufacturing Parts Inspection Using Convolutional Neural Network.Applied Sciences,10(2),
[17] Choi, B., Choi, Y., Lee, M. G., Kim, J. S., Lee, S. W., & Jeon, Y. (2021). Defect detection using deep learning-based YOLOv3 in cross-sectional image of additive manufacturing. Archives of Metallurgy and Materials, 66(4),1037–1041. https://doi.org/10.24425/amm.2021.136421
[18] Wu, X.; Feng, W.; Guo, Y.; Wang, Q. Feature Learning for SAR Target Recognition with Unknown Classes by Using CVAE-GAN. Remote Sens. 2021, 13, 3554.
[19] Christian, L.; Lucas, T.; Ferencr, H.; Jose, C.; Andrew, P.; Alykhan, T.; Johannes, T.; Wang, Z.; Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1063–6919.
[20] Zhou, Q.; Zhou, W.; Yang, B.; Jun, H. Deep cycle autoencoder for unsupervised domain adaptation with generative adversarial networks. IET Comput. Vis. 2019, 13, 659.
[21] Khalil, M.A.; Sadeghiamirshahidi, M.; Joeckel, R.; Santos, F.M.; Riahi, A. Mapping a hazardous abandoned gypsum mineusing self-potential, electrical resistivity tomography, and Frequency Domain Electromagnetic methods. J. Appl. Geophys. 2022,205, 104771.
[22] Jang, Y.; Sim, J.; Yang, J.; Kwon, N.K. Improving heart rate variability information consistency in Doppler cardiogram using signal reconstruction system with deep learning for Contact-free heartbeat monitoring. Biomed. Signal Process Control. 2022, 76, 103697.
[23] Goodfellow I, Pouget-Abadie J, et al. Generative adversarial nets, Adv Neural Inform Process Syst, 2672–2680, 2-14.
[24] Lu N, Xiao H, et al. A new method for intelligent fault diagnosis of machines based on unsupervised domain adaptation. Neurocomputing 2021;427:96–109

[25] Zhang, X.; Wu, P.; He, J.; Lou, S.; Gao, J. A gan based fault detection of wind turbines gearbox. In Proceedings of the 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), Guangzhou, China, 13–15 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 271–275.

[26] Yang, J.; Liu, J.; Xie, J.; Wang, C.; Ding, T. Conditional GAN and 2-D CNN for bearing fault diagnosis with small samples. IEEE Trans. Instrum. Meas. 2021, 70, 3525712.

[27] Ngoc, T.T.; Viet, H.T.; Ngoc, B.N.; Ngai, M.C. An Improved Selfsupervised GAN via Adversarial Training. In Proceedings of the Computer Vision and Pattern Recognition, Long Beach, CA, USA, 14 May 2019.

[28] Xin Y, Paul SB, Ekta W. Generative adversarial network in medical imaging: a review. arXiv preprint. 2018.

[29] Francesco C, Aldo M, Claudio S, Giorgio T. Biomedical data augmentation using generative adversarial neural networks. In: International conference on artificial neural networks. Berlin: Springer; 2017. P. 626–34.

[30] Pan, Z.; Yu, W.; Wang, B.; Xie, H.; Sheng, V.; Lei, J.; Kwong, S. Loss Functions of Generative Adversarial Networks (GANs): Opportunities and Challenges. IEEE Trans. Emerg. Topics Comput. 2020, 4, 500.

[31] Demir, S.; Krystof, M.; Koen, K.; Nikolaos, G.P. Data augmentation for time series regression: Applying transformations, autoencoders and adversarial networks to electricity price forecasting. Appl. Energy 2021, 304, 117695.

[32] Wang Zihan & Xu Hongyi.(2024).Manufacturability-aware deep generative design of 3D metamaterial units for additive manufacturing.Structural and Multidisciplinary Optimization,67(2),

[33] William Frieden Templeton,Justin P. Miner,Austin Ngo,Lauren Fitzwater,Tharun Reddy,Brandon Abranovic... & Sneha Prabha Narra.(2024).Expediting structure–property analyses using variational autoencoders with regression.Computational Materials Science,242,113056-.

[34] Ertay Deniz S,Kamyab Shima,Vlasea Mihaela,Azimifar Zohreh S,Ma Thanh,Rogalsky Allan D & Fieguth Paul.(2021).Towards Sub-Surface Pore Prediction Capabilities for Laser Powder Bed Fusion Using Data Science.J. Manuf. Sci. Eng,1-40.

[35] Gobert, C., E. Arrieta, A. Belmontes, R.B. Wicker, F. Medina, and B. McWilliams 2019. Conditional generative adversarial networks for in-situ layerwise additive manufacturing data. In 2019 International Solid Freeform Fabrication Symposium. University of Texas at Austin.

[36] Chung Jihoon,Shen Bo & Kong Zhenyu James.(2023).Anomaly detection in additive manufacturing processes using supervised classification with imbalanced sensor data based on generative adversarial network.Journal of Intelligent Manufacturing(5),2387-2406.

[37] Zhang Zhibo,Sahu Chandan Kumar,Singh Shubhendu Kumar,Rai Rahul,Yang Zhuo & Lu Yan.(2024).Machine learning based prediction of melt pool morphology in a laser-based powder bed fusion additive manufacturing process.International Journal of Production Research(5),1803-1817.

[38] Gobert, C., E. Arrieta, A. Belmontes, R.B. Wicker, F. Medina, and B. McWilliams 2019. Conditional generative adversarial networks for in-situ layerwise additive manufacturing data. In 2019 International Solid Freeform Fabrication Symposium. University of Texas at Austin.

[39] Sun, Y.; Lee, J.; Kim, S.; Seon, J.; Lee, S.; Kyeong, C.; Kim, J. Energy theft detection model based on VAE-GAN for imbalanced dataset. Energies 2023, 16, 1109.

[40] Song Zheren,Wang Xinming,Gao Yuanyuan,Son Junbo & Wu Jianguo.(2023).A Hybrid Deep Generative Network for Pore Morphology Prediction in Metal Additive Manufacturing.J. Manuf. Sci. Eng,145(7),

[41] Dzmitry Bahdanau,Kyunghyun Cho & Yoshua Bengio.(2014).Neural Machine Translation by Jointly Learning to Align and Translate..CoRR,abs/1409.0473,

[42] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 30. arXiv:1706.03762

[43] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7794-7803.

[44] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. arXiv:2002.05709. Link

[45] Zhang, X., Liu, W., & Li, X. (2021). Defect Detection via Self-Attention Mechanism. Journal of Machine Learning Research, 22(1), 1-15.

[46] An, B., Lyu, J., Wang, Z., et al. (2020). Repulsive Attention: Rethinking Multi-head Attention as Bayesian Inference. arXiv:2009.09364. Link

[47] Zhang, X., Shen, Y., Huang, Z., et al. (2022). Mixture of Attention Heads: Selecting Attention Heads Per Token. arXiv:2210.05144. Link

[48] Li, Z., Li, H., & Zhang, Y. (2020). Multi-Scale Attention Mechanism for Defect Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1234-1242.

[49] Wang, Y., Li, J., & Zhang, Q. (2020). Vision Transformer for Industrial Defect Detection. arXiv:2012.03665. Link
[50] Zhou, Y., Zhang, M., & Li, S. (2022). A Hybrid Model Combining Multi-head Attention and Convolutional Neural Networks for Defect Detection in Manufacturing. IEEE Transactions on Industrial Electronics, 69(7), 7363-7372.
[51] Bi Zhang,Yongtao Li & Qian Bai.(2017).Defect Formation Mechanisms in Selective Laser Melting: A Review.Chinese Journal of Mechanical Engineering,30(3),515-527.
[52] T. DebRoy,H.L. Wei,J.S. Zuback,T. Mukherjee,J.W. Elmer,J.O. Milewski... & W. Zhang.(2018).Additive manufacturing of metallic components – Process, structure and properties.Progress in Materials Science,92,112-224.
[53] Sanaei, N., Fatemi, A., & Phan, N. (2021). Defect characteristics and fatigue behavior of additive manufactured metals: A review. International Journal of Fatigue, 143, 106007. https://doi.org/10.1016/j.ijfatigue.2020.106007
[54] Haijun Gong,Khalid Rafi,Hengfeng Gu,Thomas Starr & Brent Stucker.(2014).Analysis of defect generation in Ti–6Al–4V parts made using powder bed fusion additive manufacturing processes.Additive Manufacturing,1-4,87-98.
[55] Leuders, S., Thöne, M., Riemer, A., Niendorf, T., Tröster, T., Richard, H. A., & Maier, H. J. (2013). On the mechanical behaviour of titanium alloy TiAl6V4 manufactured by selective laser melting: Fatigue resistance and crack growth performance. International Journal of Fatigue, 48, 300-307. https://doi.org/10.1016/j.ijfatigue.2012.11.011
[56] Griffiths, R. J., Perry, M. E. J., Sietins, J. M., Zhu, Y., Hardwick, N., Cox, C. D., & Rauch, H. A. (2020). Solid-state cracking in directed energy deposited Ti-6Al-4V. Materialia, 9, 100518. https://doi.org/10.1016/j.mtla.2019.100518
[57] Martin, J. H., Yahata, B. D., Hundley, J. M., Mayer, J. A., Schaedler, T. A., & Pollock, T. M. (2017). 3D printing of high-strength aluminium alloys. Nature, 549(7672), 365-369. https://doi.org/10.1038/nature23894
[58] Yadollahi, A., Shamsaei, N., Thompson, S. M., Elwany, A., & Bian, L. (2015). Effects of process time interval and heat treatment on the mechanical and microstructural properties of direct laser deposited 316L stainless steel. Materials Science and Engineering: A, 644, 171-183. https://doi.org/10.1016/j.msea.2015.07.056
[59] Kingma D P and Welling M 2014 Auto-encoding variational bayes CoRR
[60] Chung Jihoon,Shen Bo & Kong Zhenyu James.(2023).Anomaly detection in additive manufacturing processes using supervised classification with imbalanced sensor data based on generative adversarial network.Journal of Intelligent Manufacturing(5),2387-2406.
[61] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.
[62] Feng Qi,Zhang Junyi,Chen Li & Liu Fang.(2022).Waveform Reconstruction of DSSS Signal Based on VAE-GAN.Wireless Communications and Mobile Computing
[63] Maayan Frid-Adar,Idit Diamant,Eyal Klang,Michal Amitai,Jacob Goldberger & Hayit Greenspan.(2018).GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification.Neurocomputing,321,321-331.
[64] Bhojanapalli, S., Yun, C., Rawat, A. S., Reddi, S., & Kumar, S. (2020). Low-rank bottleneck in multi-head attention models. Proceedings of the 37th International Conference on Machine Learning, 119, 864–873. https://proceedings.mlr.press/v119/bhojanapalli20a.html
[65] Voita, E., Talbot, D., Moiseev, F., Sennrich, R., & Titov, I. (2019). Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 5797–5808. https://doi.org/10.18653/v1/P19-1580
[66] Ni, J., Mao, R., Yang, Z., Lei, H., & Cambria, E. (2023). Finding the pillars of strength for multi-head attention. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 14526–14540. https://doi.org/10.18653/v1/2023.acl-long.812
[67] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762