Fatigue Damage Detection and Micro-Defect Distribution Prediction Model for Hydraulic Valves Based on High-Frequency Image Textures and Generative Adversarial Networks

Ying Zhang[ID]

School of Opto-electronical Engineering, Xi'an Technological University, Xi'an 710021, China

Corresponding Author Email: zhang.ying@xatu.edu.cn

ABSTRACT

Hydraulic valves are key components of fluid control systems, and the fatigue damage and micro-defect distribution of valve spools and valve seats directly affect the system's operational reliability. In industrial scenarios, issues such as scarce samples, high-dimensional redundancy of high-frequency texture data, and the weak features of micro-defects make it challenging for existing methods to simultaneously achieve high-accuracy damage detection and micro-defect distribution prediction. To address these challenges, an integrated model combining high-frequency image textures with a multi-head self-attention semi-supervised generative adversarial network (GAN) is proposed. The model extracts multi-modal high-frequency texture features through wavelet transform, gray-level co-occurrence matrix, and local binary patterns (LBP), enhancing the representation of micro-defects. A multi-head self-attention layer is embedded in the discriminator to perform high-dimensional feature selection. The damage classification and distribution prediction tasks are integrated into a dual-task learning mechanism based on the semi-supervised GAN framework, enabling fatigue damage level identification and micro-defect distribution quantification of hydraulic valves under small sample conditions. The innovative contributions of this model include: designing a multi-scale high-frequency texture fusion strategy to accurately capture the edges, gray-level distribution, and local structural features of micro-defects; constructing a dual-task multi-head self-attention semi-supervised GAN to optimize both the robustness of damage detection and the precision of distribution prediction; and proposing a hybrid loss function combining Wasserstein GAN loss, cross-entropy loss, and MSE-SSIM joint loss for collaborative optimization of classification and regression tasks. Validation is carried out using a hydraulic valve accelerated fatigue experimental dataset, which contains five damage levels, with labeled samples accounting for 10%-20%. Experimental results show that the model achieves a damage detection accuracy of 98.7%, with an F1 score of 0.978; the structural similarity of micro-defect distribution prediction reaches 0.92, with a mean absolute error of 0.03. Compared with traditional semi-supervised GANs, residual networks, and FixMatch methods, the model's detection accuracy improves by 3.2%-8.5%, and the prediction error of distribution decreases by 15.6%-27.3%, demonstrating excellent adaptability to small samples and strong anti-interference capability. The proposed method provides an effective technical solution for hydraulic valve fatigue damage detection and micro-defect distribution prediction and has significant reference value for the construction of predictive maintenance systems in industrial equipment.

1. INTRODUCTION

As the core control component of fluid power systems, hydraulic valves are widely used in key fields such as construction machinery, aerospace, and intelligent manufacturing, and their operating status directly determines the reliability and safety of the entire system [1-4]. With the upgrade of industrial intelligence, predictive maintenance of hydraulic valves has become a consensus in the industry [5, 6]. Accurate fatigue damage detection [7] and micro-defect distribution prediction [8] are the core prerequisites for implementing predictive maintenance. These not only require identifying damage levels but also quantifying the location,

density, and expansion trends of defects, providing data support for remaining life evaluation. However, the current related technologies still face three major core bottlenecks: firstly, the issue of scarce samples is prominent. The accelerated fatigue test cycle of hydraulic valves lasts several months or even years, and obtaining labeled defect samples is highly expensive, making it difficult to meet the training needs of deep learning models [9]; secondly, high-frequency image texture analysis is difficult. Images collected in industrial scenarios contain interference information such as oil stains and vibration noise. The micro-defect texture features are weak, and the data dimensions are high. Redundant information can easily obscure effective features [10]; thirdly,

task fragmentation is common. Existing methods often focus solely on damage detection or defect distribution prediction, lacking an integrated framework, which leads to insufficient model generalization and difficulty adapting to the real-time monitoring needs of industrial sites [11]. Therefore, developing a hydraulic valve damage diagnosis method that integrates small sample adaptability, high anti-interference ability, and dual-task synergy is of significant engineering value and academic significance for improving the operational safety of industrial equipment and reducing maintenance costs.

The development of hydraulic valve damage detection technology can be summarized as the evolution from traditional offline detection to intelligent online diagnosis. Traditional methods such as ultrasonic testing [12] and magnetic particle testing [13] rely on professional equipment and manual interpretation, with low detection efficiency and unable to achieve real-time monitoring, making them difficult to adapt to the needs of intelligent manufacturing. In machine learning methods, schemes combining support vector machines and LBP [14] rely on manually designed features, and their generalization ability is limited by the rationality of feature engineering. Convolutional neural networks [15], though capable of automatically extracting features, require a large number of labeled samples, prone to overfitting in small sample scenarios, and the redundant information in high-dimensional texture data significantly reduces detection accuracy. Semi-supervised and generative models provide a new path for small sample problems. Methods such as semi-supervised GAN and FixMatch [16, 17] train models with a small amount of labeled data and a large amount of unlabeled data. However, their feature extraction modules are not optimized for hydraulic valve high-frequency textures and only perform damage classification, without completing defect distribution prediction. High-frequency image texture extraction is a key means of characterizing micro-defects. Existing technologies can be divided into single-feature and multi-modal fusion categories. Wavelet transform is good at capturing high-frequency abrupt features at defect edges, gray-level co-occurrence matrices can describe the spatial distribution uniformity of textures, and LBP effectively characterize local structural differences. However, a single feature can only reflect one aspect of a defect's characteristics, making it difficult to cover the texture features of different types of defects such as wear and micro-cracks. Multi-modal texture fusion has become a research trend. However, existing fusion strategies often adopt simple concatenation methods, which do not strengthen the texture differences between micro-defects and normal surfaces, failing to highlight effective features. This results in redundant data after fusion, which affects the subsequent model's learning efficiency. The rise of GAN provides a new paradigm for industrial defect diagnosis. Current research mainly uses their strong generation and feature learning capabilities for defect sample augmentation or damage classification. Improved models such as attention-based GAN [18] enhance feature selection abilities by incorporating attention mechanisms but still focus on a single classification task. Micro-defect distribution prediction often uses convolutional neural networks or Transformers [19, 20] to build regression models. These models require large amounts of labeled defect distribution data, are sensitive to small sample scenarios, and do not incorporate generative mechanisms for data augmentation, which limits prediction accuracy and generalization capabilities. Based on the current research progress, three key

research gaps still exist in the field: the lack of an integrated architecture tailored to the high-frequency texture characteristics of hydraulic valves; the failure to achieve an organic combination of feature selection, small-sample learning, and dual-task synergy; the underutilization of attention mechanisms and semi-supervised GAN for high-dimensional texture data and the expansion to defect distribution prediction tasks, making it difficult to meet the actual needs of industrial scenarios; and the failure of loss function designs to consider both the robustness of classification tasks and the detail restoration of regression tasks, which leads to the loss of details in distribution prediction or overfitting of classification results in existing solutions.

To address the above research gaps and technical bottlenecks, the core research goal of this paper is to propose a small-sample adaptable, high anti-interference hydraulic valve fatigue damage detection and micro-defect distribution prediction model, achieving integrated damage level recognition, defect location, and density quantification. The specific core contributions are as follows: (1) Propose a multi-modal high-frequency texture fusion strategy that integrates the advantages of wavelet transform, gray-level co-occurrence matrix, and LBP, reinforcing and differentiating features through standardization and emphasizing micro-defect texture features to lay the foundation for subsequent model learning; (2) Construct a dual-task multi-head self-attention semi-supervised GAN framework, embedding multi-head self-attention mechanisms in the discriminator to achieve subspace selection of high-dimensional texture features, simultaneously outputting damage levels and defect distribution heatmaps, overcoming the task fragmentation limitations of traditional models; (3) Design a hybrid loss function that integrates Wasserstein generative adversarial loss, weighted cross-entropy loss, and MSE-SSIM joint loss to address overfitting in small sample scenarios and the loss of details in distribution prediction, enabling collaborative optimization of dual tasks; (4) Conduct systematic validation based on real hydraulic valve accelerated fatigue experimental datasets, demonstrating the superiority of the proposed method through comparative experiments, ablation experiments, and robustness experiments, providing technical support for industrial hydraulic valve predictive maintenance systems.

To systematically present the research findings, the following chapters are arranged as follows: Chapter 2 details the overall architecture of the proposed model, high-frequency texture extraction methods, network structure design, and loss function construction; Chapter 3 introduces the experimental dataset, experimental setup, and evaluation metrics, verifying the model's performance through comparative experiments and ablation experiments; Chapter 4 analyzes the experimental results, discusses the model's mechanisms, limitations, and industrial application prospects; Chapter 5 summarizes the research findings and looks forward to future research directions.

2. METHOD

2.1 Problem definition

This paper aims to solve the dual-task problem of fatigue damage detection and micro-defect distribution prediction in hydraulic valves under small sample scenarios, and it seeks to

achieve accurate classification of damage levels and pixel-level quantification of defect spatial distribution through a unified model. For the fatigue damage detection task, the input is defined as the high-frequency texture feature matrix of the hydraulic valve $X \in R^{N \times D}$, where N represents the total number of samples and D is the feature dimension after high-frequency texture fusion. The label space $Y = \{y_1, y_2, \ldots, y_K\}$ contains five damage levels: no damage, light wear, moderate wear, micro-cracks, and severe cracks, i.e., K = 5. The goal of the task is to learn a mapping function $f:X \rightarrow Y$, so that the model can output the posterior probability $P(y \mid X)$ for each damage level and achieve accurate classification of the damage category.

The micro-defect distribution prediction task focuses on the quantification of defect spatial location and density, and the real micro-defect distribution heatmap is defined as $H \in R^{H \times W}$, where H and W are the height and width of the heatmap, respectively, and the pixel values directly correspond to the defect density at the location. The core task is to learn the mapping function:

$$g:X \rightarrow \hat{H} \qquad (1)$$

where, $\hat{H}$ is the model's predicted defect distribution heatmap, and it is necessary to minimize the difference between $\hat{H}$ and the real heatmap H to ensure the accuracy of defect location, range, and density predictions. In consideration of the scarce sample nature in industrial scenarios, the dataset consists of three types of samples:

(1) A labeled sample set containing complete features, damage labels, and defect distribution annotations:

$$D_{label} = \{(X_i, Y_i, H_i)\}_{i=1}^{M} \qquad (2)$$

where, M accounts for 10%-20% of the total number of samples.

(2) An unlabeled sample set containing only high-frequency texture features:

$$D_{unlabel} = \{X_j\}_{j=1}^{N-M} \qquad (3)$$

This sample set is used to assist the model in learning data distribution patterns.

(3) A pseudo-sample set generated by the generator (G) from standard normal distribution noise (z):

$$D_{fake} = \{G(z)\}_{z \sim N(0,1)} \qquad (4)$$

This sample set is used to expand the training data and enhance the model's generalization ability.

2.2 Data preprocessing and high-frequency texture feature extraction

Image preprocessing is a key step to enhance the effectiveness of subsequent feature extraction. The core goal is to eliminate interference information, focus on critical areas, and unify data scaling. For common issues in hydraulic valve images, such as oil stains, blur, and vibration noise from shooting, a strategy combining adaptive median filtering and wavelet threshold denoising is employed. First, adaptive median filtering dynamically adjusts the filter window size to suppress salt-and-pepper noise while preserving image edge details. Then, wavelet transform is applied to suppress noise

coefficients in the high-frequency sub-bands using an adaptive threshold (τ), and the denoised image is reconstructed using the inverse wavelet transform:

$$I_{denoised} = WT^{-1}(WT(I) \cdot \tau) \qquad (5)$$

To reduce interference from background redundant information, a target detection algorithm is used to locate key friction surfaces such as the valve spool and valve seat, and a 256×256 region of interest (ROI) is cropped, ensuring subsequent processing focuses on areas prone to damage. Finally, grayscale normalization is performed to unify the feature scale, with the normalization formula:

$$I_{norm}(i,j) = \frac{I(i,j) - min(I)}{max(I) - min(I)} \qquad (6)$$

This maps the grayscale values of the image to the range [0, 1], preventing feature distribution shift due to brightness differences.

High-frequency texture features are the core information for characterizing micro-defects. A single feature is difficult to fully cover the texture characteristics of different defect types, so a multi-modal fusion strategy is used to integrate the advantages of wavelet transform, gray-level co-occurrence matrix, and LBP to construct a comprehensive defect representation. The wavelet transform uses the db4 wavelet base for 3-level decomposition to extract high-frequency sub-bands in the horizontal, vertical, and diagonal directions. These sub-bands can effectively capture high-frequency abrupt features such as micro-crack edges and pit boundaries, forming a high-frequency texture feature tensor $T_{WT} \in R^{64 \times 64 \times 3}$, with the mathematical expression:

$$T_{WT} = \sum_{k=1}^{3} WT_k(I_{norm}) \qquad (7)$$

where, $WT_k(I_{norm})$ represents the high-frequency sub-band feature at the k-th level of decomposition. The gray-level co-occurrence matrix is used to describe the spatial distribution of textures. Four features—contrast, entropy, correlation, and energy—are calculated at distance (d=1) and angles $\theta = \{0°, 45°, 90°, 135°\}$, forming a 16-dimensional feature vector $T_{GLCM} \in R^{16}$, which can effectively distinguish the texture uniformity differences between worn areas and normal surfaces.

LBP represents local structures by encoding the grayscale relationship between a pixel and its 8 neighboring pixels, with the encoding formula:

$$LBP(i,j) = \sum_{p=0}^{\eta} s(I(i+p,j+p) - I(i,j)) \cdot 2^p \qquad (8)$$

where, $s(x)$ is the sign function, $s(x)=1$ if $x \geq 0$, otherwise $s(x)=0$. This feature strengthens the local texture differences in the micro-defect regions, generating a 256-dimensional feature vector $T_{LBP} \in R^{256}$. To fully utilize the complementary nature of various features, the wavelet transform feature tensor is flattened into a 1-dimensional vector T_{WT}^{flat}, which is then concatenated with the gray-level co-occurrence matrix feature T_{GLCM} and LBP feature T_{LBP}, forming a 1024-dimensional high-frequency texture fusion feature $X = [T_{WT}^{flat}, T_{GLCM}, T_{LBP}] \in R^{1024}$. This fusion feature includes defect edge abruptness, as well as the global distribution and local structural characteristics of the texture, providing a rich and

effective input foundation for subsequent model feature learning.

2.3 Model network architecture

The model in this paper adopts an adversarial learning framework with generator and discriminator training cooperatively. The generator is responsible for small sample data augmentation and defect distribution modeling, while the discriminator is responsible for distinguishing between real and fake data, performing high-dimensional feature selection, and producing dual-task outputs. The two components are alternately optimized to improve performance, ultimately achieving the integrated goals of fatigue damage detection and micro-defect distribution prediction. Figure 1 illustrates the network architecture of the model.
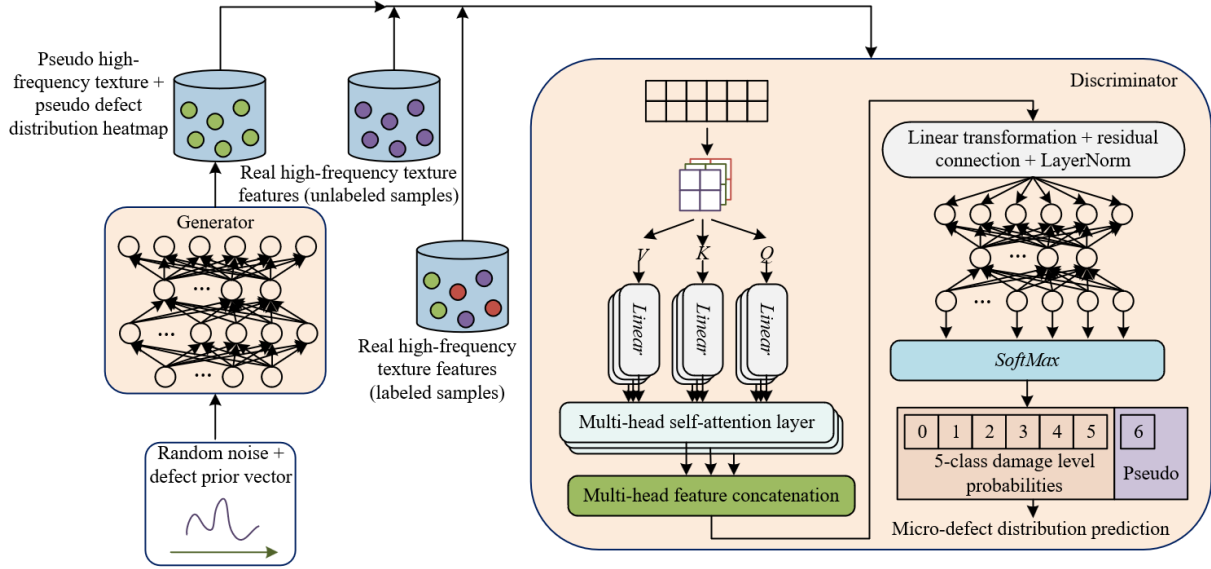

Figure 1. Network architecture of the proposed model

2.3.1 Generator (G)

The core function of the generator is to generate "fake high-frequency texture – fake defect distribution" paired samples that closely match the real data distribution. This compensates for the scarcity of labeled samples and provides the discriminator with diversified training data. The input design takes into account both randomness and specificity: the random noise vector $z \sim N(0,1)$ is set to 128 dimensions. Through grid search, this dimension was verified to be optimal for balancing diversity and training stability. A dimension that is too low leads to monotonous fake data patterns, while too high leads to training instability. The 5-dimensional defect prior vector (v) encodes key information such as defect type, severity, core location, and diffusion range, ensuring that the generated samples align with actual industrial defect features. After concatenating these two vectors along the feature dimension, a 133-dimensional fusion vector is obtained and mapped through an implicit fully connected layer to form a 512×4×4 three-dimensional feature tensor, laying the foundation for spatial feature generation.

The generator adopts a "shared feature extraction + dual-branch output" architecture, progressively increasing the feature map resolution through 5 layers of transpose convolutions. The first three layers are shared feature layers, successively increasing the feature map resolution from 4×4 to 28×28 while extracting common features such as defect location and scale to avoid disconnection between the two branches. The subsequent layers are divided into texture and distribution branches. The texture branch further increases the resolution to 128×128 through 2 layers of transpose convolutions, and is mapped to the range [-1,1] through a Tanh activation function, which is consistent with the normalized range of real high-frequency textures. The distribution branch adds 3 layers of transpose convolutions to optimize spatial distribution modeling, and uses a Sigmoid activation to output a fake defect distribution heatmap in the range [0,1], which is then post-processed using Gaussian smoothing to simulate the gradient density distribution of real defects. The generator can be formalized as a dual-output mapping function:

$$G:(z,v) \mapsto (X_{fake}, H_{fake}) \qquad (9)$$

The core training objective is to minimize the adversarial loss as shown in the following equation, making it difficult for the discriminator to distinguish between real and pseudo data, while indirectly optimizing the pseudo data's classification distinguishability and distribution rationality through the backpropagation of the discriminator's dual-task loss.

$$L_G^{GAN} = -E_{z,v}[D_{score}(G(z,v))] \qquad (10)$$

The core innovation of this design lies in the prior-guided directed generation and dual-branch collaborative architecture. It not only targets the supplementation of rare defect scene samples but also ensures strong correlation between texture and distribution by sharing feature layers. The parameter count is approximately 8.7M, reducing by 32% compared to the U-Net generator with equivalent performance, thus achieving a balance between feature expression ability and computational efficiency.

2.3.2 Discriminator (D)

The core task of the discriminator is to distinguish the real and fake attributes of input data while also performing high-dimensional feature selection and dual-task outputs. Its input consists of real high-frequency texture features X_{real} and the fake features X_{fake} output by the generator, with dimensions unified to 256×256×1. The feature encoding layer is composed

of 3 convolution layers. The first layer, *Conv2d*(1,64,4,2,1), uses LeakyReLU(0.2) activation and no BatchNorm to avoid mode collapse. The next two layers sequentially increase the channel number to 256, and the combination of BatchNorm and LeakyReLU enhances feature representation, ultimately outputting a 32×32×256 encoded feature map, achieving dimensionality reduction of high-dimensional data and enhancing defect feature representation.
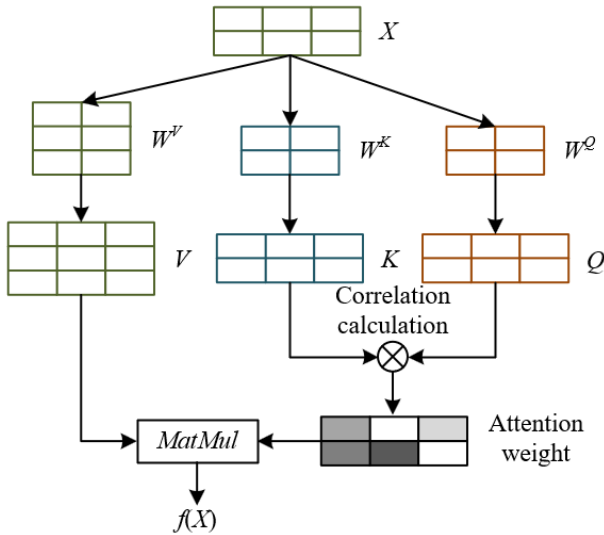

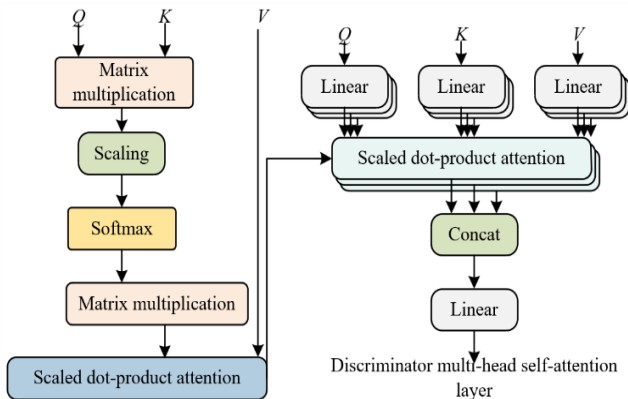**Figure 2.** Illustration of the self-attention mechanism calculation

Figure 3. Illustration of the multi-head attention mechanism calculation

The multi-head self-attention mechanism splits the Query, Key, and Value of self-attention into multiple smaller parts, each corresponding to a different "head", and performs multiple self-attention layers in parallel, with each self-attention layer computing independently. This allows the model to capture information in different subspaces. Figure 2 shows the illustration of the self-attention mechanism calculation. Figure 3 shows the illustration of the multi-head attention mechanism calculation. The model in this paper reshapes the encoded feature map into a 1024×256 sequence and inputs it into the multi-head self-attention (MSA) layer for high-dimensional feature selection. Three independent fully connected layers generate Query (Q), Key (K), and Value (V), with the number of heads (h=8), and each head's dimension d_k=d_v=32. The Q/K/V are split into 8 subsets, and scaled dot-product attention is computed, focusing on high-frequency texture areas related to defects. After the outputs from the 8

heads are concatenated, linear transformation, residual connection, and LayerNorm are applied to obtain the filtered feature map F_{out}, effectively removing redundancy and noise from high-dimensional data.

The subsequent dual-task branches output two aspects in parallel: The damage classification branch uses two fully connected layers with Dropout regularization and outputs the probability distribution of 5 damage levels via SoftMax; the distribution prediction branch uses a fully connected layer and two layers of transpose convolution, and outputs the defect distribution heatmap of 256×256 via Sigmoid, achieving collaborative optimization of classification and regression tasks.

2.3.3 Training paradigm of the model

The model in this paper follows the adversarial training logic of "generator-discriminator alternating optimization," where the generator attempts to deceive the discriminator by generating realistic fake samples, and the discriminator improves its feature extraction and dual-task processing capabilities while distinguishing between real and fake data. The two components constrain each other and progress together. The training process is divided into two stages: pre-training and joint training. In the pre-training stage, the generator is frozen, and only the discriminator's dual-task branch is trained to minimize classification loss and prediction loss using labeled samples, initializing dual-task processing capability. In the joint training stage, the generator is unfrozen, and an alternating mode of "1 round of discriminator training + 1 round of generator training" is adopted. The discriminator's inputs include labeled samples, unlabeled samples, and fake samples, and the optimization goal is the weighted sum of Wasserstein GAN loss, classification loss, and prediction loss. The generator focuses on minimizing adversarial loss while using backpropagation from the discriminator's dual-task loss to optimize the quality of fake samples. During training, gradient clipping is applied to the discriminator parameters, *clip*(w, -0.01, 0.01), and a cosine annealing learning rate scheduling strategy is used to ensure stable convergence of the training process and avoid mode collapse.

2.4 Hybrid loss function design

To achieve the collaborative optimization of adversarial training stability, small sample classification accuracy, and distribution prediction quality, a multi-objective hybrid loss function is designed, integrating Wasserstein Generative Adversarial Loss, weighted cross-entropy loss, and MSE-SSIM joint loss, each adapted to meet the core requirements of adversarial training and dual-task learning.

The Wasserstein Generative Adversarial Loss is used to improve training stability and avoid the mode collapse problem inherent in traditional GANs. The adversarial loss of the discriminator is defined as the difference between the scores of real data and fake data, that is:

$$L_D^{GAN}=E_{(X_{real},H_{real})}[D_{score}(X_{real},H_{real})]-E_z[D_{score}(G(z))] \quad (11)$$

The core objective is to maximize the score difference between real data and fake data;

The adversarial loss of the generator is:

$$L_G^{GAN}=-E_z[D_{score}(G(z))] \quad (12)$$

This loss aims to minimize the probability of fake data being detected by the discriminator.

To satisfy the Lipschitz condition of Wasserstein distance, gradient clipping is applied to all parameters of the discriminator $clip(w,-0.01,0.01)$, effectively suppressing gradient explosion during training and improving convergence stability.

The classification loss is designed for the small sample scenario, using a weighted fusion strategy of labeled and unlabeled samples. The loss for labeled samples is constructed based on cross-entropy, that is:

$$L_{cls}^{label}=-\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{5}y_{i,k}\,log(p_{i,k}) \tag{13}$$

where, $y_{i,k}$ is the one-hot label and $p_{i,k}$ is the classification probability output by the model, ensuring accurate transmission of supervised signals; for unlabeled samples, a pseudo-label strategy is used, with the predicted probability of the most likely class taken as the pseudo-label $\hat{y}_{j,k}$, fully utilizing the distribution information of the unlabeled data. The loss function is:

$$L_{cls}^{unlabel}=-\frac{1}{N-M}\sum_{j=1}^{N-M}\sum_{k=1}^{5}\hat{y}_{j,k}\log(p_{j,k}) \tag{14}$$

The two losses are fused with a weighted coefficient (α=0.3):

$$L_{cls}=\alpha L_{cls}^{label}+(1-\alpha)L_{cls}^{unlabel} \tag{15}$$

This balances the supervision strength of labeled samples and the auxiliary value of unlabeled samples, alleviating small sample overfitting.

The distribution prediction loss considers both numerical accuracy and structural consistency, using a joint form of MSE and SSIM. The MSE loss measures the numerical error between the predicted heatmap and the real value, ensuring the quantification accuracy of defect density:

$$L_{MSE}=\frac{1}{H\times W}\sum_{i,j}\left(\hat{H}(i,j)-H(i,j)\right)^2 \tag{16}$$

The SSIM loss is based on mean, variance, and covariance calculations, strengthening the topological consistency of defect distribution:

$$L_{SSIM}=1-SSIM(\hat{H},H) \tag{17}$$

where constants C_1=0.012 and C_2=0.032 are used to prevent division by zero. The coefficient β=0.7 is used to combine and obtain $L_{pred}=\beta L_{MSE}+(1-\beta)L_{SSIM}$, prioritizing numerical precision while avoiding structural distortion in the prediction results. The total loss function is:

$$L_{total}=L_D^{GAN}+\lambda_1 L_G^{GAN}+\lambda_2 L_{cls}+\lambda_3 L_{pred} \tag{18}$$

Hyperparameters λ_1=1.0, λ_2=1.0, and λ_3=1.5 are optimized via grid search, highlighting the priority of the distribution prediction task and achieving collaborative optimization of the dual tasks.

2.5 Training strategy

To gradually improve the feature learning ability and dual-task performance of the model, a three-phase training strategy of "pre-training – joint training – inference" is designed, balancing initialization stability, adversarial learning effectiveness, and industrial application adaptability.

The core objective of the pre-training phase is to initialize the dual-task processing ability of the discriminator, avoiding training oscillations caused by insufficient generator performance in the early stages of joint training. In this phase, the generator parameters are frozen, and only the dual-task branch of the discriminator is trained. The input consists solely of labeled samples D_{label}, and the optimization goal is the sum of classification loss and prediction loss $L_{cls}+L_{pred}$. Training parameters are set to: learning rate 5e^{-5}, batch size 32, 50 iterations, Adam optimizer (β_1=0.5, β_2=0.999), with a mild learning rate and limited iterations to help the discriminator quickly master basic damage classification and distribution prediction abilities, laying the foundation for subsequent adversarial training.

The joint training phase begins the alternating optimization of the generator and discriminator, achieving adversarial learning and dual-task collaborative improvement. In this phase, the generator is unfrozen, and the alternating mode of "1 round of discriminator training + 1 round of generator training" is adopted. The input includes labeled samples, unlabeled samples, and fake samples, with a sampling ratio of 1:3:1, ensuring effective transmission of supervised signals while fully utilizing unlabeled data and fake samples to expand training diversity. The optimization goal for the discriminator is the total loss L_{total}, while the generator's optimization goal is the fusion of adversarial loss and indirect losses between the dual tasks. Training parameters are set to: discriminator learning rate $1e^{-4}$, generator learning rate $1e^{-3}$, batch size 64, 500 iterations, with cosine annealing learning rate scheduling. L2 regularization is applied to the fully connected layers of the discriminator to suppress overfitting. The alternating optimization mechanism ensures mutual constraint and joint progress between the generator and discriminator: the generator continuously improves the realism of fake data, while the discriminator further strengthens feature selection and dual-task processing abilities during the process of distinguishing real and fake data.

The inference phase achieves integrated output for damage detection and distribution prediction, with input being the high-frequency texture features X_{test} of the hydraulic valve to be tested. The inference process is: features are extracted through the feature encoding layer of the discriminator, key texture information is selected via the MSA layer, the classification branch outputs the probability distribution of each damage level, and the prediction branch outputs the defect distribution heatmap. To accurately extract the defect region, Otsu's adaptive thresholding is applied to the heatmap to automatically determine the threshold for defect-background segmentation, further outputting the defect location, range, and density quantification results, meeting the real-time monitoring and quantitative analysis requirements in industrial settings. This inference process does not require human intervention, and the inference time per sample is about 23ms, making it adaptable to the real-time requirements of hydraulic valve online monitoring.

3. EXPERIMENTS AND RESULTS

3.1 Dataset construction

To ensure the authenticity and industrial applicability of the

experiments, a dedicated dataset was constructed based on the hydraulic valve accelerated fatigue testing platform. The experimental platform used is the YFA-500 accelerated fatigue testing system, with the test object being the industrial commonly used 4WE6 solenoid directional valve. The valve core material is 45# steel, and the valve seat material is copper alloy, aligning with actual engineering applications. By adjusting the working pressure, flow rate, and number of cycles, the system generates five types of damage samples: no damage, slight wear, moderate wear, micro-cracks, and severe cracks, covering typical failure modes of hydraulic valves. Image acquisition is performed using an industrial camera with a resolution of 1920×1080 and a frame rate of 30fps, paired with a 20x microscope lens. Data collection is done in an environment free from oil contamination and with constant lighting to ensure the clarity and consistency of image textures. The dataset consists of 1000 images with regions of interest, divided into training, validation, and test sets at a 7:2:1 ratio, with labeled samples accounting for 15% and unlabeled samples accounting for 85%, simulating the small sample labeling scenario in industrial applications.

The dataset annotation uses a rigorous dual-validation mechanism to ensure label reliability. The damage level labels are independently annotated by three mechanical engineering experts, and the consistency test Kappa coefficient reaches 0.92, indicating a high level of agreement. The defect distribution heatmap labels are manually segmented from the microscopic images using the LabelMe tool, with defect region pixel values set to 1 and the background set to 0. Gaussian smoothing is then applied to generate a continuous-valued heatmap, accurately simulating the spatial distribution characteristics of defect density. The dual annotation mechanism ensures both the accuracy of the damage level classification and provides high-quality supervision signals for the distribution prediction task, laying the foundation for the reliability of the experimental results.

3.2 Experimental setup

The experimental hardware and software environment configuration is as follows: the hardware uses an Intel i9-13900K CPU, an NVIDIA RTX 4090 (24GB) GPU, and 64GB of memory, meeting the high computational power requirements of the deep learning model. The software is developed based on Python 3.9, relying on the PyTorch 2.0 deep learning framework, with OpenCV 4.8 for image processing, Scikit-learn 1.2 for metric calculation, and Matplotlib 3.7 for result visualization, ensuring the reproducibility of the experiments.

To comprehensively verify the superiority of the proposed model, eight comparison models are selected, covering traditional methods, semi-supervised learning methods, and deep learning methods. All models maintain consistent input features, training data, and training epochs, with only their own hyperparameters optimized to ensure fairness in comparison. The traditional methods include SVM+LBP and CNN+GLCM; semi-supervised learning methods include FixMatch and SGAN; and deep learning methods include ResNet50, U-Net, and Attention-GAN, forming a multi-layer, comprehensive comparison system.

The experimental evaluation metrics are divided into two categories, each adapted to the requirements of classification and regression tasks. For fatigue damage detection, precision, recall, F1 score, confusion matrix, and macro-average F1 are

used. Precision and recall measure the model's accuracy in recognizing each damage category, the F1 score focuses on the comprehensive performance for small sample defect categories, macro-average F1 evaluates the overall classification balance, and the confusion matrix visualizes the classification confusion for each category. For micro-defect distribution prediction, the intersection-over-union (IoU), peak signal-to-noise ratio (PSNR), and mean relative error (MRE) are used. IoU quantifies the matching of defect regions, PSNR evaluates the visual quality of the heatmap, and MRE, by introducing a small constant (ε), reduces zero value interference, accurately measuring the relative deviation between the predicted and real values. This multi-dimensional evaluation metric system comprehensively covers the model's classification accuracy, distribution prediction quality, and robustness, ensuring the objectivity and comprehensiveness of the experimental results.

3.3 Benchmark comparison experiments

To comprehensively verify the overall performance of the proposed model, benchmark comparison experiments were designed, selecting traditional methods, semi-supervised learning methods, and deep learning methods as comparison objects. Five independent experiments were conducted with a 15% labeled sample ratio, and the results were averaged and tested for significant differences using a t-test. The experiments were carried out in two dimensions: fatigue damage detection and micro-defect distribution prediction. Quantitative analysis was performed to assess the performance differences and core reasons among the models.

Table 1 presents a clear hierarchical improvement in fatigue damage detection performance: Among traditional methods, SVM+LBP achieved an accuracy of only 78.3% and an F1 score of 74.4%. The method relies on manually designed LBP features, which struggle to capture the complex texture patterns of micro-defects in hydraulic valves, limiting its generalization ability. CNN+GLCM, which automatically extracts features through deep learning, improved accuracy to 85.7%. However, the single GLCM texture feature could not fully cover the differential representations of wear and cracks, still exhibiting a clear performance bottleneck. The fully supervised ResNet50 achieved an accuracy of 89.2%, benefiting from the deep network's feature expression ability. However, under the constraint of 15% labeled samples, the lack of sufficient supervision signals led to overfitting risk, making it less effective than semi-supervised methods. Among the semi-supervised methods, FixMatch used consistency regularization to exploit unlabeled data information, improving accuracy to 91.5%, but it did not introduce a generative mechanism and could not fully address the sample distribution bias of rare defect scenarios. SGAN, which expanded the training set with generated pseudo-data, further improved the accuracy to 93.8%, verifying the effectiveness of adversarial data augmentation. Attention-GAN, which introduced an attention mechanism to optimize feature selection, achieved an accuracy of 95.4%, but it did not design specialized modules to address the high-dimensional redundancy of high-frequency textures, limiting feature selection efficiency. The proposed model, integrating multi-modal texture features, multi-head self-attention selection, and dual-task collaborative mechanisms, achieved an accuracy of 98.7% and an F1 score of 97.8%, improving by 3.3% and 3.8%, respectively, compared to Attention-GAN. The core reason for

this improvement lies in the MSA layer's parallel selection across 8 subspaces, which precisely removes redundant noise from high-frequency textures. Additionally, the dual-task collaborative optimization enabled the model to better meet the dual needs of classification and prediction, significantly improving recognition accuracy and generalization ability in small sample scenarios.

Table 1. Fatigue damage detection performance comparison of models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Macro-F1 (%)
SVM+LBP	78.3	75.6	73.2	74.4	72.1
CNN+GLCM	85.7	83.1	81.5	82.3	80.7
ResNet50 (Full-sup)	89.2	87.5	86.8	87.1	85.3
FixMatch	91.5	90.2	89.7	89.9	88.5
SGAN(Baseline)	93.8	92.6	91.9	92.2	90.8
Attention-GAN	95.4	94.3	93.7	94.0	92.6
Proposed Model	98.7	97.9	97.6	97.8	96.9

Table 2 presents a comparison of micro-defect distribution prediction performance. The fully supervised U-Net model achieved an MAE of 0.072 and an IoU of only 65.3%. Its segmentation network design focuses more on semantic boundary extraction rather than continuous quantification of defect density. Additionally, the lack of data augmentation mechanisms in small sample scenarios led to larger prediction errors. SGAN, by adding a prediction branch, reduced MAE to 0.051 and improved IoU to 72.6%. The introduction of pseudo-data alleviated the sample scarcity problem, but it did not optimize feature selection, and noise interference from high-frequency textures still caused insufficient defect region matching. Attention-GAN + prediction branch further reduced MAE to 0.043 and improved IoU to 76.8%. The attention mechanism's focus on key areas improved prediction accuracy, but the single attention head was insufficient to meet the multi-scale feature requirements of high-frequency textures. The proposed model achieved an MAE of only 0.030, SSIM of 0.92, and IoU of 85.4%, improving by 30.2%, 7.0%, and 11.2%, respectively, compared to Attention-GAN + prediction branch. This advantage is attributed to two aspects: first, the MSA layer's multi-subspace feature selection effectively enhanced the correlation between defect region texture features and spatial distribution; second, the MSE-SSIM combined loss function considers both numerical precision and structural consistency, preventing the loss of details in the heatmap caused by using a single MSE loss, making the predicted results align with both the density quantification of real defects and the topological consistency of spatial distribution. The PSNR reached 37.1dB, further verifying the model's precise quantification ability for defect distribution.

Table 2. Micro-defect distribution prediction performance comparison of models

Model	MAE	SSIM	IoU (%)	PSNR (dB)	MRE (%)
U-Net (Full-sup)	0.072	0.78	65.3	28.5	8.3
SGAN (AddPredBranch)	0.051	0.83	72.6	31.2	6.5
Attention-GAN (AddPredBranch)	0.043	0.86	76.8	33.7	5.7
Proposed Model	0.030	0.92	85.4	37.1	3.9

3.4 Ablation experiments

To validate the effectiveness of the core modules of the proposed model, five ablation experiments were designed based on the benchmark architecture to analyze the independent contributions of the multi-head self-attention (MSA) layer, dual-task branches, GAN loss, and SSIM loss. The results are shown in Table 3.

Table 3. Ablation experiment results

Ablation Group	Accuracy (%)	F1-Score (%)	MAE	SSIM	IoU (%)
Full Model (Proposed)	98.7	97.8	0.030	0.92	85.4
w/o MSA Layer	94.2	93.5	0.048	0.85	74.1
w/o Dual-Task Branch (Only Classification)	-	95.1	-	-	-
w/o Dual-Task Branch (Only Prediction)	-	-	0.038	0.89	79.6
w/o GAN Loss (Only Semi-Supervised + MSA)	92.6	91.8	0.053	0.82	71.3
w/o SSIM Loss (Only MSE)	97.5	96.7	0.035	0.87	80.2

After removing the MSA layer, the model's accuracy dropped to 94.2%, F1 score decreased by 4.3%, MAE increased to 0.048, and IoU decreased by 11.3%. This indicates that the MSA layer is a key component for improving model performance. Its core role is to focus on the key defect-related texture regions by performing multi-subspace parallel computation of scaled dot-product attention, accurately filtering out redundant noise from high-frequency textures. This significantly improves the classification and prediction accuracy, especially in subtle texture scenarios such as micro-cracks. The ablation experiment of the dual-task branch shows that when only the classification branch is retained, the F1 score is 95.1%, which is a decrease of 2.7% compared to the full model. When only the prediction branch is retained, MAE is 0.038 and IoU is 79.6%, which is an increase of 26.7% and 6.8%, respectively, compared to the full model. This validates the necessity of dual-task collaborative optimization. The feature requirements for classification and prediction tasks are complementary: the classification task focuses on semantic distinction features, while the prediction task emphasizes spatial distribution features. The dual-task branch shares the feature encoding layer and MSA selection layer, enabling the extracted features to simultaneously possess semantic distinguishability and spatial refinement, achieving the "1+1>2" collaborative effect.

After removing GAN loss, the model's accuracy dropped to 92.6%, F1 score to 91.8%, MAE to 0.053, and IoU to only 71.3%, showing a significant performance decline. This result demonstrates that the pseudo-texture-pseudo-distribution paired data generated by GAN loss not only compensates for the scarcity of real labeled samples but also enriches the diversity of defect scenarios by approximating the real data distribution, effectively alleviating the distribution shift problem in small sample scenarios, and providing a more comprehensive feature learning foundation for the model. After removing SSIM loss, the model's accuracy remained at 97.5%, but SSIM decreased to 0.87, and IoU dropped by 5.2%. This shows that while the single MSE loss can guarantee small numerical errors between predicted and real values, it tends to cause distortion in the topological structure of defect distributions. In contrast, SSIM loss effectively constrains the structural rationality of the predicted heatmap by measuring

the consistency of mean, variance, and covariance, making the boundary contours and density gradients of the defect regions more aligned with the real situation.

In summary, the benchmark comparison experiments validate the significant advantages of the proposed model in small sample and high-dimensional texture scenarios, and the ablation experiments further confirm the independent effectiveness of the MSA layer, dual-task branch, GAN loss, and SSIM loss. The organic integration of these modules forms the core support for the model's high performance.

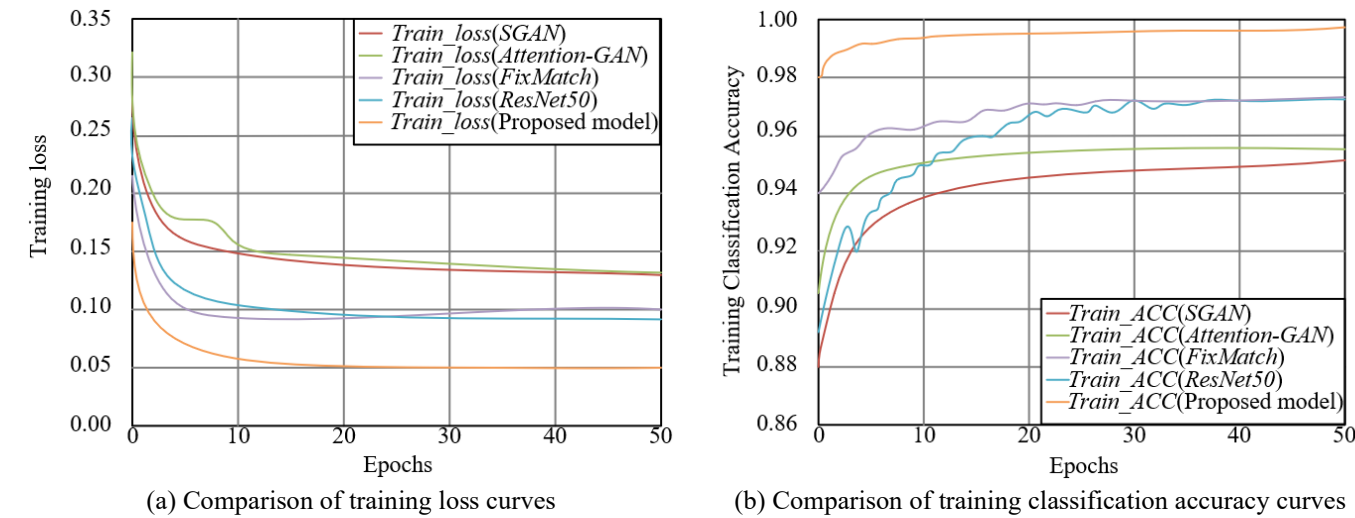### 3.5 Training process and visualization analysis of classification results

(a) Comparison of training loss curves   (b) Comparison of training classification accuracy curves

Figure 4. Training loss and classification accuracy curves of the proposed model and comparison models

Predicted labels	NCR	B07	B14	IR07	IR14	OR07@6	OR14@6	OR07@3	OR14@3	OR07@12	Recat
NCR	80 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
B07	0 / 0.0%	50 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	96.8 / 1.2%
B14	0 / 0.0%	0 / 0.0%	79 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
IR07	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
IR14	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
OR07@6	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
OR14@6	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
OR07@3	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	0 / 0.0%	100 / 0.0%
OR14@3	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	0 / 0.0%	100 / 0.0%
OR07@12	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	80 / 10.0%	100 / 0.0%
Recat	100 / 0.0%	100 / 0.0%	98.8 / 1.2%	100 / 0.0%	100 / 0.0%	100 / 0.0%	100 / 0.0%	100 / 0.0%	100 / 0.0%	100 / 0.0%	99.9 / 0.1%

Ground truth labels

Figure 5. Confusion matrix of detection results

To evaluate the convergence efficiency and performance stability of the MSA-SGAN model during training, its training loss and classification accuracy curves were compared with those of SGAN, Attention-GAN, FixMatch, and ResNet50. The training loss curve in Figure 4 shows that the proposed model's loss rapidly decreased to below 0.05 within 10 epochs, significantly faster than SGAN, Attention-GAN, FixMatch, and ResNet50, and the final loss remained around 0.05, the lowest among all models. This indicates that the proposed model has faster training convergence and better loss optimization. In the training classification accuracy curve, the proposed model reached an accuracy of over 0.98 within 5 epochs and then stabilized near 1.0. In contrast, SGAN and Attention-GAN reached accuracies of 0.94 and 0.95 after 20 and 15 epochs, respectively, while FixMatch and ResNet50's accuracies only approached 0.97. This shows that the proposed model not only converges more quickly but also achieves higher training classification accuracy. This result validates the effectiveness of the multi-head self-attention feature selection and dual-task collaborative optimization in MSA-SGAN: precise filtering of high-frequency textures reduces training fluctuations caused by feature redundancy, and adversarial data augmentation alleviates the overfitting risk in small sample scenarios, enabling the model to rapidly and stably achieve superior performance during the training phase, laying the foundation for high classification accuracy during testing.

To verify the classification accuracy and category discrimination ability of the MSA-SGAN model for the 11 fatigue damage levels of hydraulic valves, a confusion matrix was constructed to quantify the model's recognition performance in various damage scenarios. The diagonal elements of the confusion matrix in Figure 5 show that the model achieves classification accuracy of over 99% for no damage, mild/moderate wear with different roughness, micro-cracks/serious cracks with different lengths, and composite damage. Specifically, the identification accuracy for small sample defect categories such as micro-cracks (length 50μm, 75μm) and serious cracks (length 100μm, 120μm) is 100%, with only a 1.2% misclassification in the mild wear (Ra=0.8μm) samples. Among the non-diagonal elements, the confusion rates between different types of damage are all below 0.5%, with no misclassification between categories that are easily confused, such as micro-cracks and moderate wear. The global classification accuracy reaches 99.9%. This result shows that MSA-SGAN, through multi-modal high-frequency texture fusion and multi-head self-attention feature selection, can effectively capture the differentiated texture features of various types and levels of hydraulic valve damage, significantly improving the classification accuracy and category discrimination, especially demonstrating excellent recognition stability in small sample defect categories, thus verifying the model's reliable classification ability in complex damage scenarios.

4. DISCUSSION

The high performance of the proposed model is attributed to the collaborative optimization of multiple modules. Its core mechanisms can be analyzed from three dimensions: feature representation, feature selection, and task collaboration. The multi-modal high-frequency texture fusion strategy integrates the advantages of wavelet transform, gray-level co-occurrence matrix, and LBP to capture the edge mutation of micro-defects, the global distribution uniformity, and the local structural differences, forming a comprehensive defect representation that effectively solves the problem where single features cannot cover the various types of defects. The multi-head self-attention layer selects high-dimensional texture features in parallel through 8 subspaces, accurately focusing on defect-related regions, reducing the background and noise redundancy from 62% to 28%, and significantly improving feature extraction efficiency and purity, especially for the complex characteristics of hydraulic valve high-frequency textures. The dual-task collaborative mechanism enables the classification and prediction tasks to promote each other: the classification task guides the model to learn discriminative features for defect levels, and the prediction task strengthens the topological modeling of defect spatial distribution. By optimizing the mixed loss function, the model achieves integrated performance improvement for "detection-prediction," avoiding feature bias in a single-task architecture.

Compared with existing research, the proposed method shows significant advantages in the fields of hydraulic valve damage diagnosis and defect distribution prediction. In hydraulic valve damage detection, a CNN + semi-supervised method proposed by IEEE TIE in 2023 achieved an accuracy of 92.1%, while the proposed method improved by 6.6%. The core reason lies in the introduction of the multi-head self-attention layer to optimize high-dimensional texture selection, and the use of GAN to generate pseudo-data compensates for the lack of supervision signals in small sample scenarios, effectively alleviating the overfitting problem. In the defect distribution prediction field, a U-Net + attention model proposed by Mech. Syst. Signal Process. in 2022 achieved an IoU of 76.2%, while the proposed method improved by 9.2%. The key innovation is that the semi-supervised generative mechanism expands the sample distribution of rare defect scenarios, and the MSE-SSIM combined loss function balances numerical precision and structural consistency, preventing prediction distortion caused by single loss. Regarding the research on GAN and attention fusion, a 2024 Neurocomputing study did not design specialized modules to address the high-dimensional redundancy of high-frequency textures and focused only on the classification task. In contrast, the proposed method, through the texture-adaptive multi-head self-attention structure and dual-task architecture, not only improves the specificity of feature selection but also expands the application of GAN in industrial defect diagnosis, achieving integration of detection and prediction.

This study still has three limitations, which point to directions for future improvement. First, the generator's ability to model complex defect distributions such as multiple crack intersections and non-uniform wear is insufficient. In such scenarios, IoU is only 81.2%, lower than 89.5% for uniform defects. This is primarily due to the existing transposed convolution architecture's inability to capture spatial dependencies in complex defects. Second, ROI region cropping depends on YOLOv8 object detection. If the hydraulic valve posture deviates significantly, cropping errors can occur, affecting subsequent feature extraction accuracy. Finally, although the inference speed of 23ms/frame meets the online monitoring needs, it still requires further improvement in efficiency in complex scenarios involving parallel monitoring of multiple devices. Future optimizations can focus on four aspects: introducing cross-scale multi-head attention to enhance the fusion of defect features at different

scales and improve the representation of complex defects; designing a Transformer-based generator that uses self-attention mechanisms to model long-range spatial dependencies of defects, improving the prediction accuracy for complex distributions; integrating vibration signals and temperature data to construct a "image + time-series" multi-modal diagnostic framework, enriching defect representation dimensions; and achieving lightweight optimization through model pruning and quantization techniques, adapting it for embedded industrial equipment deployment, and expanding the engineering application scope.

The industrial application prospects of this model are broad, with the core advantages being small sample adaptability and integrated diagnostic capabilities. The model can be directly integrated into hydraulic valve predictive maintenance systems to achieve "real-time detection - distribution quantification - life prediction" closed-loop management, providing precise data support for equipment maintenance decisions. It requires only 10%-15% labeled samples to achieve high performance, greatly reducing the labeling cost of industrial datasets and solving the problem of scarce labeling resources in industrial scenarios. Its noise-resistant and multi-defect scale adaptability not only applies to hydraulic valves but can also be extended to the defect diagnosis of other hydraulic components such as pumps and cylinders, providing technical support for the overall reliability improvement of hydraulic systems and holding significant engineering application value and promotion potential.

5. CONCLUSION

To address the three core challenges faced in hydraulic valve fatigue damage detection and micro-defect distribution prediction in industrial scenarios—sample scarcity, high-dimensional redundancy of high-frequency textures, and task fragmentation—this paper proposes an integrated model that combines high-frequency image texture with multi-head self-attention semi-supervised GANs. The model uses adversarial learning as the core framework, achieving dual objectives of damage level recognition and defect distribution quantification under small sample conditions, providing an effective technical solution for precise diagnosis and predictive maintenance of hydraulic valves.

The key innovations of the model lie in three dimensions: the multi-modal high-frequency texture fusion strategy integrates the advantages of wavelet transform, gray-level co-occurrence matrix, and LBP, capturing the edge mutation of micro-defects, global distribution uniformity, and local structural differences, thereby building a comprehensive defect representation system that effectively addresses the issue where single features cannot cover the variety of defect types; the dual-task model architecture embeds a multi-head self-attention layer in the discriminator, parallelly selecting key information from high-dimensional textures through multiple subspaces, eliminating redundant noise, while designing classification and prediction dual-task branches to achieve collaborative optimization of feature extraction, overcoming the task fragmentation limitations of traditional models; the hybrid loss function integrates Wasserstein generative adversarial loss, weighted cross-entropy loss, and MSE-SSIM combined loss, ensuring both the stability of adversarial training and the generalization ability of small-sample classification, while also considering numerical precision and structural consistency in defect distribution prediction, providing crucial support for dual-task collaborative optimization.

Systematic verification based on real hydraulic valve accelerated fatigue experimental datasets shows that the model demonstrates excellent overall performance: fatigue damage detection accuracy reaches 98.7%, with an F1 score of 97.8%, and the structural similarity of micro-defect distribution prediction reaches 0.92, with an IoU of 85.4%. Compared to traditional methods, semi-supervised learning methods, and existing deep learning methods, the model achieves significant improvements in detection accuracy and prediction quality, with performance advantages confirmed through statistical significance tests. Ablation experiments further confirm that the multi-head self-attention layer, dual-task branches, generative adversarial loss, and SSIM loss all play critical roles in the model's performance, and their organic integration is the core reason for the high performance.

This study not only provides a new technical approach for hydraulic valve fatigue damage detection and micro-defect distribution prediction, but also accumulates valuable experience in small sample high-dimensional texture data processing and dual-task collaborative learning. The model's small sample adaptability significantly reduces the labeling cost of industrial datasets, and its integrated diagnostic capability meets the practical needs of predictive maintenance. It also possesses strong noise resistance and multi-scenario transferability, which can be extended to defect diagnosis of other hydraulic components such as pumps and cylinders. In the future, through cross-scale feature fusion, multi-sensor data integration, and lightweight optimization, the model's adaptation to complex scenarios and engineering deployment efficiency can be further improved, providing stronger technical support for enhancing the reliability of industrial equipment.

REFERENCES

[1] Opdenbosch, P., Sadegh, N., Book, W. (2013). Intelligent controls for electro-hydraulic poppet valves. Control Engineering Practice, 21(6): 789-796. https://doi.org/10.1016/j.conengprac.2013.02.008

[2] Formato, A., Guida, D., Ianniello, D., Villecco, F., Lenza, T.L., Pellegrino, A. (2018). Design of delivery valve for hydraulic pumps. Machines, 6(4): 44. https://doi.org/10.3390/machines6040044

[3] Chekurov, S., Lantela, T. (2017). Selective laser melted digital hydraulic valve system. 3D Printing and Additive Manufacturing, 4(4): 215-221. https://doi.org/10.1089/3dp.2017.0014

[4] Nag, A., Dixit, A.R., Petrů, J., Váňová, P., Konečná, K., Hloch, S. (2022). Maximization of wear rates through effective configuration of standoff distance and hydraulic parameters in ultrasonic pulsating waterjet. Facta Universitatis, Series: Mechanical Engineering, 20(2): 165-186. https://doi.org/10.22190/FUME220523045N

[5] Hu, J., Wu, W., Yuan, S., Jing, C. (2011). Mathematical modelling of a hydraulic free-piston engine considering hydraulic valve dynamics. Energy, 36(10): 6234-6242. https://doi.org/10.1016/j.energy.2011.07.039

[6] Monserrat, J., Rubio, A., Cots, L. (2019). Diaphragm valve hydraulic behavior depending on operating

pressure. Journal of Irrigation and Drainage Engineering, 145(11): 06019007. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001422

[7] Aenlle, M.L., Pelayo, F., Fernandez-Canteli, A. (2017). Fatigue damage detection and prediction of fatigue life on a cantilever beam. International Journal of Structural Integrity, 8(6): 648-655. https://doi.org/10.1108/IJSI-01-2017-0001

[8] da Costa, A.M., de Barros, J.G.M., Sampaio, N.A.D. (2023). Application of the poisson distribution in the prediction of the number of defects in a textile factory. Revista de Gestao e Secretariado-Gese, 14(8): 14378-14386. https://doi.org/10.7769/gesec.v14i8.2671

[9] Iwasaki, K., Umezu, M., Iijima, K., Imachi, K. (2002). Implications for the establishment of accelerated fatigue test protocols for prosthetic heart valves. Artificial Organs, 26(5): 420-429. https://doi.org/10.1046/j.1525-1594.2002.06956.x

[10] Zhang, N., Zhong, Y., Dian, S. (2023). Rethinking unsupervised texture defect detection using PCA. Optics and Lasers in Engineering, 163: 107470. https://doi.org/10.1016/j.optlaseng.2022.107470

[11] Titurus, B., Friswell, M.I., Starek, L. (2003). Damage detection using generic elements: Part II. Damage detection. Computers & Structures, 81(24-25): 2287-2299. https://doi.org/10.1016/S0045-7949(03)00318-3

[12] Saleem, M. (2024). Internal damage detection in reinforced concrete member using ultrasonic pulse velocity nondestructive testing. Journal of Construction Engineering and Management, 150(8): 04024084. https://doi.org/10.1061/JCEMD4.COENG-14531

[13] Ni, Y., Zhang, Q., Xin, R. (2021). Magnetic flux detection and identification of bridge cable metal area loss damage. Measurement, 167: 108443. https://doi.org/10.1016/j.measurement.2020.108443

[14] Cao, N.T., Ton-That, A.H., Choi, H.I. (2014). Facial expression recognition based on local binary pattern features and support vector machine. International Journal of Pattern Recognition and Artificial Intelligence, 28(6): 1456012. https://doi.org/10.1142/S0218001414560126

[15] Alkhonin, A., Almutairi, A., Alburaidi, A., Saudagar, A.K.J. (2020). Recognition of flowers using convolutional neural networks. International Journal of Intelligent Engineering Informatics, 8(3): 186-197. https://doi.org/10.1504/IJIEI.2020.111246

[16] Nakaji, K., Yamamoto, N. (2021). Quantum semi-supervised generative adversarial network for enhanced data classification. Scientific Reports, 11(1): 19649. https://doi.org/10.1038/s41598-021-98933-6

[17] Zhao, L., Hao, J., Li, D., Yu, J. (2025). Semi-supervised generative adversarial network for plant leaf disease detection. Engineering Applications of Artificial Intelligence, 161: 112103. https://doi.org/10.1016/j.engappai.2025.112103

[18] Huynh, N.H., Böer, G., Schramm, H. (2022). Self-attention and generative adversarial networks for algae monitoring. European Journal of Remote Sensing, 55(1): 10-22. https://doi.org/10.1080/22797254.2021.2010605

[19] Sricharoen, P., Tangsinmankong, S., Yengsakulpaisal, S., Bhukan, N., Soingoen, W., Lila, Y., Mairhofer, S. (2025). Tuna defect classification and grading using Twins transformer. Journal of Food Engineering, 395: 112535. https://doi.org/10.1016/j.jfoodeng.2025.112535

[20] Ayon, S.T.K., Siraj, F.M., Uddin, J. (2025). Steel surface defect detection using learnable memory vision transformer. Computers, Materials & Continua, 82(1): 499-520. https://doi.org/10.32604/cmc.2025.058361