IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 6, December, 2025, pp. 3453-3468
Journal homepage: http://iieta.org/journals/ts

Image Processing Algorithms Combining Transfer Learning and Semi-Supervised Learning for Industrial Part Defect Detection

Dongliang Zhang[1,2*], Xiaoli Li[3]

[1] School of Information Engineering and Artificial Intelligence, Henan Open University, Zhengzhou 450046, China
[2] School of Information Engineering and Artificial Intelligence, Zhengzhou Vocational College of Information and Technology, Zhengzhou 450046, China
[3] Puyang Institute of Technology, Henan University, Puyang 457000, China

Corresponding Author Email: zhangdongliang@haou.edu.cn

https://doi.org/10.18280/ts.420632

Received: 22 May 2025
Revised: 8 November 2025
Accepted: 22 November 2025
Available online: 31 December 2025

Keywords:
industrial defect detection, transfer learning, semi-supervised learning, meta-learning, cross-domain adaptation, expected risk minimization, self-supervised prior mining, small-sample learning

ABSTRACT

In the context of intelligent manufacturing, industrial part defect detection is a core process in quality control, facing three key challenges: high labeling costs leading to small sample constraints, cross-domain distribution shifts caused by differences in production lines and equipment, and low feature distinguishability due to the similarity between small or rare defects and background textures. Existing methods combining transfer learning and semi-supervised learning are often engineering-driven approaches, which face theoretical limitations such as negative transfer, broad generalization error boundaries, and insufficient utilization of mutual information from unlabeled data. To address these challenges, this paper proposes a new paradigm for cross-domain semi-supervised expected risk minimization driven by meta-learning, and constructs a meta-adaptive defect detection framework (MADD-Framework). This framework achieves collaborative optimization of domain-invariant feature learning, pseudo-label noise suppression, and normal mode prior mining at the theoretical level. The framework consists of three core components: First, a multi-view teacher-student network (MV-TSN) integrating domain-adaptive adapters to mitigate cross-domain distribution shifts through domain-specific data augmentation and domain-invariant feature consistency constraints. Second, a reconstruction contrastive self-supervision module that narrows theoretical generalization error boundaries by modeling the normal mode prior of industrial parts. Third, a meta-adaptive pseudo-label optimization module (MAPOM) that integrates domain difference statistics to achieve dynamic threshold adjustment and pseudo-label purification through dual-layer optimization, enhancing the utilization efficiency of unlabeled data. This framework breaks through the limitations of current methods combining transfer learning and semi-supervised learning, providing a new theoretical paradigm and engineering solution for small-sample, cross-domain defect detection in intelligent manufacturing, contributing to the industrial implementation of intelligent defect detection technology.

1. INTRODUCTION

The deepening advancement of Smart Manufacturing 4.0 imposes strict requirements on the accuracy, efficiency, and environmental adaptability of industrial part quality inspection [1-3]. The surface and internal defects of industrial parts directly affect the safety of end products [4, 5], potentially leading to serious hazards and losses. Traditional manual inspection suffers from low efficiency, high subjectivity, and high missed detection rates. Although deep learning-based methods have improved accuracy, they rely on large-scale labeled data, making them difficult to adapt to the reality of scarce defect samples and high labeling costs in industrial scenarios [6, 7]. Currently, industrial part defect detection faces three core bottlenecks: small sample constraints caused by diverse defect types with low incidence rates, with labeling costs exceeding the capacity of small and medium enterprises [8]; cross-domain adaptation challenges due to production line, equipment, and environmental differences causing data distribution shifts, where models trained in a single domain lack generalization ability [9, 10]; and tiny and rare defects with weak features that are highly similar to background textures, which are easily disturbed by noise and difficult to effectively identify [11, 12]. Therefore, constructing a fusion algorithm of transfer learning and semi-supervised learning that is both theoretically rigorous and practically useful is essential to fundamentally alleviate these bottlenecks. It will play a crucial theoretical and practical role in promoting the industrial implementation of intelligent inspection technology, reducing quality control costs, and improving the level of smart manufacturing.

Transfer learning is the core technology for solving cross-domain defect detection. Mainstream methods can be divided into feature alignment and domain-adaptive network approaches. The former minimizes the domain feature distribution difference through metrics like maximum mean discrepancy (MMD) and correlation alignment [13, 14], while the latter reduces domain shifts through specialized network structures [15]. However, existing methods often adopt static alignment strategies, which are unable to cope with dynamic distribution changes in industrial scenarios, and the fusion with semi-supervised learning is generally limited to a shallow mode of "pre-training + fine-tuning," which can lead to negative transfer and insufficient adaptation to the special characteristics of industrial defects. This paper proposes a domain-adaptive adapter that implements dynamic domain adaptation in the training process of transfer learning and semi-supervised learning. By jointly optimizing domain alignment loss and semi-supervised consistency loss, the issue of negative transfer is fundamentally alleviated. Semi-supervised learning provides an effective solution for small sample problems. Consistency regularization and pseudo-label purification are current hotspots, but existing methods rely on manually designed augmentation strategies that fail to simulate the real variations of industrial scenarios. Pseudo-label purification overlooks domain differences, leading to noise accumulation, and does not fully utilize the mutual information of unlabeled data. To address these shortcomings, this paper designs an industrial-specific multi-view enhancement strategy to accurately simulate real detection variations. An MAPOM integrates domain difference statistics to achieve domain-adaptive purification and enhance the utilization efficiency of unlabeled data.

To address both small sample and cross-domain challenges simultaneously, current fusion methods of transfer learning [16, 17] and semi-supervised learning [18, 19] often adopt a module concatenation approach. These methods initialize models with transfer learning and then optimize performance with semi-supervised learning, but such approaches lack a unified theoretical framework. The modules have poor synergy, with domain adaptation and semi-supervised training being independent of each other, and there is a lack of industry-specific regularization constraints. This results in insufficient generalization ability and defect recognition sensitivity. This paper proposes a new paradigm for cross-domain semi-supervised expected risk minimization driven by meta-learning, unifying domain adaptation from transfer learning, unlabeled data utilization from semi-supervised learning, and prior mining from self-supervised learning under the expected risk minimization framework, achieving deep collaboration among the three. Additionally, self-supervised learning tasks such as reconstruction and contrastive learning are mostly designed for general vision tasks and lack industrial part structure specificity. Meta-learning's dynamic threshold adjustment and memory enhancement methods do not consider cross-domain differences, making them difficult to adapt to complex distribution changes in industrial scenarios. The reconstruction contrastive self-supervised module (RCSM) in this paper is designed based on the characteristics of industrial parts, mining the normal mode prior, while the MAPOM incorporates domain difference statistics into the meta-learner to achieve cross-domain adaptive dynamic optimization, improving model robustness.

Current research still faces key issues: fusion methods lack a unified theoretical framework, making it difficult to alleviate negative transfer and generalization errors; engineering designs lack specificity to industrial scenarios [20, 21], failing to meet the requirements of scene adaptation, cross-domain robustness, and weak feature sensitivity; pseudo-label optimization overlooks domain differences, leading to training instability. To address these issues, this paper constructs a fusion framework with both theoretical depth and practical value. The core contributions are: proposing a meta-learning-driven cross-domain semi-supervised expected risk minimization paradigm, deriving the generalization error boundary, and providing unified theoretical support for the deep fusion of the two; designing an MV-TSN with integrated domain-adaptive adapters to alleviate negative transfer and improve cross-domain robustness through exclusive enhancement and dynamic domain alignment; proposing a RCSM to mine the normal mode prior and convert it into dynamic anomaly attention maps, enhancing the ability to identify tiny and rare defects; constructing a MAPOM that integrates domain difference statistics, achieving pseudo-label purification with dual-layer optimization and domain-adaptive memory, ensuring training stability; verifying effectiveness through multi-dimensional experiments on three international benchmark datasets and two custom industrial datasets, and proposing a lightweight solution that achieves real-time inference at 32 frames per second on embedded devices to meet industrial deployment needs.

The subsequent sections are arranged as follows: Section 2 elaborates on the MADD-Framework, including problem definition, theoretical formalization, network design, mathematical derivation, and training algorithms; Section 3 validates the method's effectiveness through performance comparison, ablation experiments, statistical tests, and failure case analysis; Section 4 discusses core findings, theoretical and practical significance, limitations, and future directions; Section 5 concludes the paper.

2. METHODS

2.1 Problem definition

This paper studies the problem of small-sample cross-domain defect detection in the context of industrial part inspection, and first clarifies the definitions of source and target domains and the boundaries of the core task. The source domain consists of labeled samples from publicly available industrial defect datasets:

$$D_s=\{(x_s^i,y_s^i)\}_{i=1}^{N_s} \tag{1}$$

where, N_s is the total number of labeled samples in the source domain, x_s^i represents the source domain image sample, and $y_s^i\in\{0,1,...,C\}$ is the corresponding class label. 0 denotes a normal sample, and 1 to C correspond to different types of defect samples. The target domain D_t includes both small-labeled data $D_t^l=\{(x_t^j,y_t^j)\}_{j=1}^{N_l}$ and large unlabeled data $D_t^u=\{x_t^k\}_{k=1}^{N_u}$, where $N_l \ll N_s$ and $N_u \gg N_l$, which reflects the scarcity of labeled resources in industrial scenarios. From the perspective of domain distribution, the feature distributions of the source and target domains satisfy $P_s(x)\neq P_t(x)$, i.e., there exists a cross-domain distribution shift, while the conditional class distributions satisfy $Q_s(y|x)=Q_t(y|x)$, i.e., the label mapping given the features remains consistent between the two domains. The core task of this paper is to learn a robust defect detector $f{:}x_t\rightarrow(y_t,b_t)$, where y_t is the defect class prediction for the target domain sample, and b_t is the bounding box coordinates of the defect region, ultimately achieving high accuracy and robustness in detecting target domain samples.

To ensure the rationality and feasibility of the method, three key assumptions based on the actual characteristics of industrial defect detection are proposed. First, the source and target domains share core defect features, meaning that the essential structural characteristics of defects such as cracks, dents, and wear remain consistent across the domains. The difference between domains is reflected only in the statistical distribution of the data, which provides a theoretical foundation for knowledge transfer in transfer learning. Second, the proportion of normal samples in the unlabeled data of the target domain is significantly higher than that of defect samples, and it covers various defect types. This corresponds to the reality in industrial production, where qualified products dominate and defects appear sporadically. This assumption provides a premise for semi-supervised learning to utilize unlabeled data to augment supervisory information. Third, the normal mode of industrial parts exhibits strong structural regularity, such as the regular geometric shapes of mechanical parts and the uniform texture distribution of welds. These inherent patterns can be effectively mined through self-supervised learning tasks, providing prior knowledge support for distinguishing between defect and normal samples. These assumptions are based on actual observations from industrial scenarios and provide clear constraints and foundations for the subsequent method design.

2.2 Theoretical motivation and formalization

The core framework of traditional transfer learning and semi-supervised learning fusion methods is expected risk minimization, and its risk function can be represented as:

$$R(f)=\lambda R_l(f)+(1-\lambda)R_u(f)+\Omega(f) \tag{2}$$

where, $R_l(f)=E_{(x,y)\sim D_s\cup D_t}[\mathrm{L}(f(x),y)]$ is the labeled risk, representing the prediction error of the model on labeled samples from both the source and target domains; $R_u(f)=E_{x\sim P_i^u}[\mathrm{L}(f(x),\hat{y})]$ is the unlabeled risk, optimizing the model with pseudo-labels $\hat{y}$ using unlabeled data; $\Omega(f)$ is a general regularization term used to constrain model complexity; λ is the coefficient that balances the labeled and unlabeled risks. However, this paradigm has significant limitations in the small-sample cross-domain industrial part defect detection scenario: the cross-domain distribution shift leads to systematic biases in the estimation of labeled and unlabeled risks. The general regularization term does not fully utilize the structural prior knowledge of industrial parts, and the quality of pseudo-labels is prone to noise accumulation due to domain shifts, all of which result in poor model generalization performance.

To address the shortcomings of the traditional paradigm, this paper proposes a meta-learning-driven cross-domain semi-supervised expected risk minimization paradigm and constructs an improved risk function:

$$R_{meta\text{-}adapt}(f)=\lambda R_l^{\alpha}(f)+(1-\lambda)R_u^{p}(f)+\alpha\mathrm{R}_{da}(f)+\beta\mathrm{R}_{prior}(f) \tag{3}$$

where, $R_l^{\alpha}(f)=E_{(x,y)\sim D_s\cup D_t}[A(x)\cdot\mathrm{L}(f(x),y)]$ is the labeled risk with domain adaptation weights, $A(x)$ is the output of the domain-adaptive adapter, and the loss weights of samples from different domains are dynamically adjusted to alleviate cross-domain bias; $R_u^{p}(f)=E_{x\sim D_t}[P(x)\cdot\mathrm{L}(f(x),\hat{y}_{\tau(x)})]$ is the unlabeled risk with pseudo-label quality weights, $P(x)$ is the confidence weight of the pseudo-label, and $\tau(z)$ is the dynamic threshold predicted by the meta-learner based on domain difference features (z), ensuring the quality of the pseudo-labels; $\mathrm{R}_{da}(f)=Dist(P_s(f(x)),P_t(f(x)))$ is the domain alignment risk, which suppresses cross-domain shifts by measuring and minimizing the feature distribution difference between the source and target domains; $\mathrm{R}_{prior}(f)=E_{x\sim D_{normal}}[\mathrm{L}_{rec}(f_{dec}(f(x)),x)]$ is the prior regularization risk, which constrains the model based on reconstruction errors of normal samples, mining the normal mode prior of industrial parts.

Based on PAC-Bayes theory, this paper further derives the generalization error bounds of the new paradigm to verify its theoretical soundness. The boundary expression is:

$$R_{meta\text{-}adapt}(f)=\lambda R_t^{a}(f)+(1-\lambda)R_u^{p}(f)+\alpha\mathrm{R}_{da}(f)+\beta\mathrm{R}_{prior}(f) \tag{4}$$

where, $R_{true}(f)$ is the true risk of the model on the real data distribution of the target domain, f_0 is the initial model constructed from the pre-trained backbone network, and $KL(f\|f_0)$ is the Kullback-Leibler divergence between the current model and the initial model, used to measure the deviation of model parameters; δ is the confidence level, and $N=N_s+N_l+N_u$ is the total number of labeled and unlabeled samples in the source and target domains; γ is the domain shift coefficient, quantifying the impact of cross-domain distribution differences on generalization performance. This boundary clearly shows that the proposed domain alignment risk $R_{da}(f)$ can directly reduce $Dist(P_s,P_t)$, and the prior regularization risk $R_{prior}(f)$ reduces model complexity by mining industrial priors, thereby reducing $KL(f\|f_0)$. The synergistic effect of both reduces the generalization error boundary, providing a solid theoretical basis for the design of subsequent modules.

2.3 Overall framework design

The proposed MADD-Framework is centered around cross-domain semi-supervised expected risk minimization. It achieves the mapping of the theoretical paradigm to engineering practice through four collaborative modules, where the key components of the risk function correspond to the functions of each module, forming a closed-loop optimization system. The overall structure of the framework is shown in Figure 1, clearly presenting the data flow transmission path and the loss collaboration mechanism: The Transfer Initialization Module (TIM) performs domain-adaptive initialization of the model based on source domain knowledge and small labeled target domain samples, providing a robust parameter starting point for subsequent learning. The MV-TSN generates differential samples through industrial-specific data augmentation, combining domain-invariant feature consistency constraints and dynamic domain alignment loss to effectively suppress cross-domain shifts, corresponding to the domain alignment risk term in the risk function. The RCSM mines the normal mode prior based on the structural characteristics of industrial parts, transforming the reconstruction error into dynamic anomaly attention maps, which enhance defect region feature identification and reduce model generalization error through prior regularization risk terms. The MAPOM integrates domain difference statistical information, implementing dual-layer optimization to adjust dynamic thresholds and purify pseudo-labels, ensuring the effective utilization of unlabeled data, corresponding to the unlabeled risk term with pseudo-label quality weight. The four modules work in close collaboration, jointly optimizing and minimizing $R_{meta\text{-}adapt}(f)$, ultimately achieving high-precision detection of industrial part defects in small-sample cross-domain scenarios.

2.4 TIM

The TIM adopts a two-stage transfer learning strategy to provide a robust model parameter starting point for subsequent cross-domain semi-supervised training. The first stage realizes the transfer of general visual features to industrial domain features by selecting ResNet-50 or ViT-Base as the backbone network. These networks have been pre-trained on the ImageNet dataset for general visual feature extraction, providing strong feature extraction capabilities. Based on this, fine-tuning is performed using source domain industrial defect data, enabling the network to learn the general discriminative features between defects and normal samples in industrial scenarios, thus adapting to the image characteristics of industrial parts. The second stage completes the transfer adaptation from the source domain to the target domain by inserting a domain-adaptive adapter into the higher layers of the backbone network. This layer is chosen to balance feature abstraction and domain specificity, effectively adjusting the cross-domain feature distribution difference. The structure of the domain-adaptive adapter is defined as:

$$DAA(f)=LayerNorm(W_2 \cdot ReLU(W_1 \cdot f+b_1)+b_2+f) \quad (5)$$

where, f is the high-dimensional feature output by the backbone network, W_1 and W_2 are the weight matrices of the two fully connected layers of the adapter, and b_1 and b_2 are the corresponding bias terms. The residual connection design avoids feature degradation, ensuring the transmission of original valid features. During training, only the parameters of the domain-adaptive adapter and the top-level parameters of the backbone network are fine-tuned, while the lower-level parameters are frozen to retain the representation ability of general visual features, thus alleviating negative transfer in cross-domain migration from the parameter update perspective.

To further narrow the feature distribution difference between the source and target domains, the module introduces dynamic domain alignment loss, using the CORAL loss function to align the distribution of high-dimensional features. This loss is particularly suitable for the high-dimensional, complex defect features in industrial scenarios, as it can effectively measure and minimize the second-order statistical difference of features between the two domains. The loss function is defined as:

$$L_{da}=\frac{1}{4d^2}\|Cov(f_s)\text{-}Cov(f_t)\|_F^2 \quad (6)$$

where, $Cov(f_s)$ and $Cov(f_t)$ represent the covariance matrices of the source and target domain features, describing the second-order statistical distribution of features; d is the feature dimension, and $\| \|_F{}^2$ is the Frobenius norm used to compute the matrix difference. The dynamic domain alignment loss is not optimized independently but participates in the overall optimization process in conjunction with the loss functions of subsequent modules. By measuring the feature distribution difference between the two domains in real-time and backpropagating the gradients, it drives the domain-adaptive adapter to dynamically adjust the feature mapping, thus achieving dynamic alignment of feature distributions between the source and target domains and reducing the negative impact of cross-domain shifts on subsequent semi-supervised learning. Figure 1 shows the framework structure of the TIM.

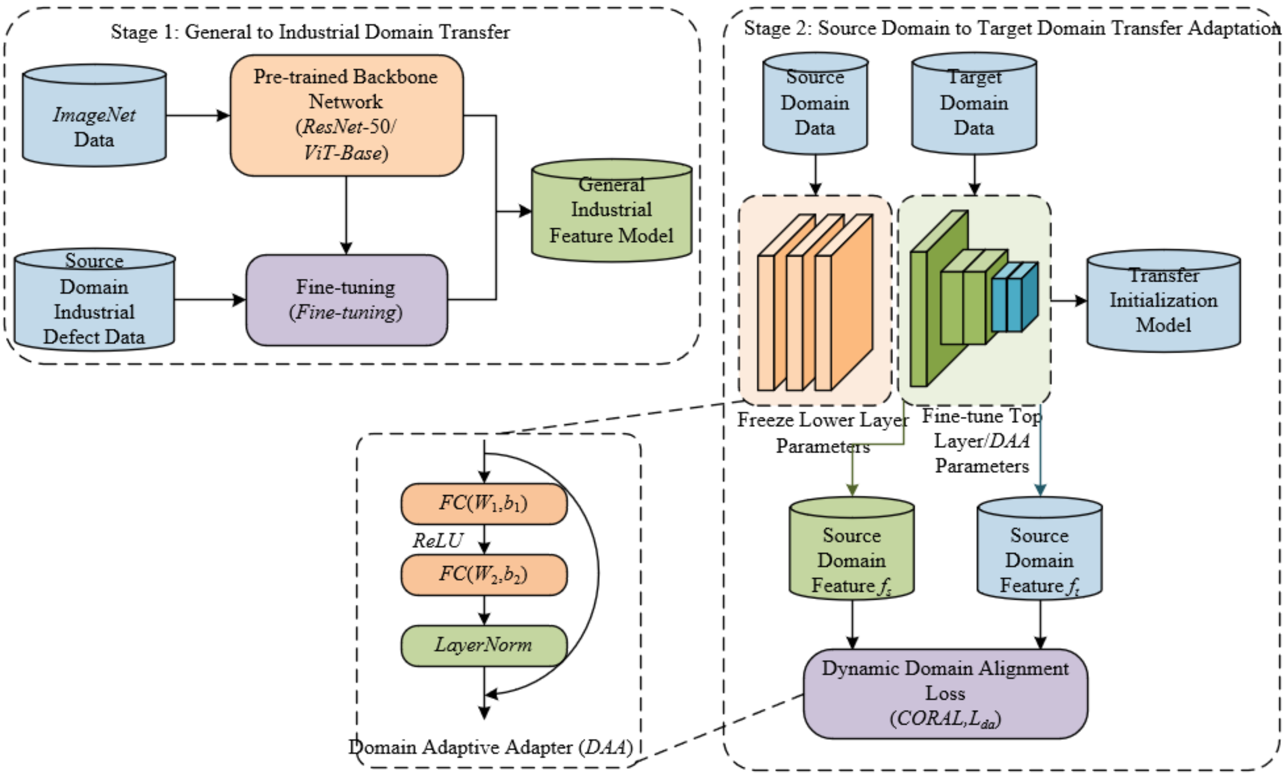

Figure 1. Framework structure of the TIM

2.5 MV-TSN

The MV-TSN adopts a "one teacher, two students" architecture, consisting of a teacher network and two structurally identical student networks. The three networks share a common backbone network and domain-adaptive adapter for basic feature extraction, ensuring consistency and synergy in feature extraction capabilities. The network architecture is shown in Figure 2. The teacher network is initialized by the TIM and does not participate in gradient updates during training. Every 5 training cycles, the weights of student network A are copied to the teacher network, which uses its relatively stable parameters to generate high-confidence pseudo-labels, providing reliable supervision signals for semi-supervised learning on unlabeled data. Both student networks are constructed based on the basic feature extraction architecture, with the core difference being their input data augmentation strategies. Multi-view augmentation specific to industrial scenarios generates samples with distribution differences, driving the model to learn domain-invariant features. Student network A adopts a mild augmentation strategy, simulating slight environmental fluctuations in the same detection equipment through brightness adjustment and low-intensity Gaussian noise addition. Student network B adopts an intensive augmentation strategy, simulating real data variations caused by cross-device and cross-angle detection through affine transformations, contrast reversal, and industry-specific noise injection. The noise types are optimized for the imaging methods, with salt-and-pepper noise for ultrasound images and artifact interference for CT images, ensuring the industrial scene adaptability of the augmented samples.

To strengthen the model's robustness to industrial variations and enhance domain-invariant feature learning, the network introduces domain-invariant feature consistency loss, which includes prediction consistency loss and feature consistency loss. Prediction consistency loss minimizes the output difference between the two student networks, constraining the model's insensitivity to non-essential variations. The loss function is defined as:

$$L_{cons\text{-}p}=MSE(p_A(x_t),p_B(x_t^a)) \tag{7}$$

where, x_t is the original target domain sample, x_t^a is the heavily augmented sample from student network B, and p_A and p_B are the predicted class probability distributions from the two student networks. The mean squared error (MSE) is used to measure the difference in the predicted distributions. Feature consistency loss uses MMD to align the domain-adaptive features of the two student networks, forcing the model to extract core features independent of augmentation methods. The loss function is defined as:

$$L_{cons\text{-}f}=\left\|\frac{1}{N}\sum_{i=1}^{N}\phi\,(f_A^{daa,i})-\frac{1}{N}\sum_{i=1}^{N}\phi\,(f_B^{daa,i})\right\|_2^2 \tag{8}$$

where, f_A^{daa} and f_B^{daa} are the domain-adaptive features output by the two student networks, and $\phi(\ )$ represents the mapping to the Reproducing Kernel Hilbert Space (RKHS). The mean difference of the mapped features is computed to align the distribution of high-dimensional feature spaces. These two types of losses work synergistically, constraining the model at both the output prediction and intermediate feature levels, effectively enhancing the representation ability of domain-invariant features and the model's robustness to industrial scene variations.

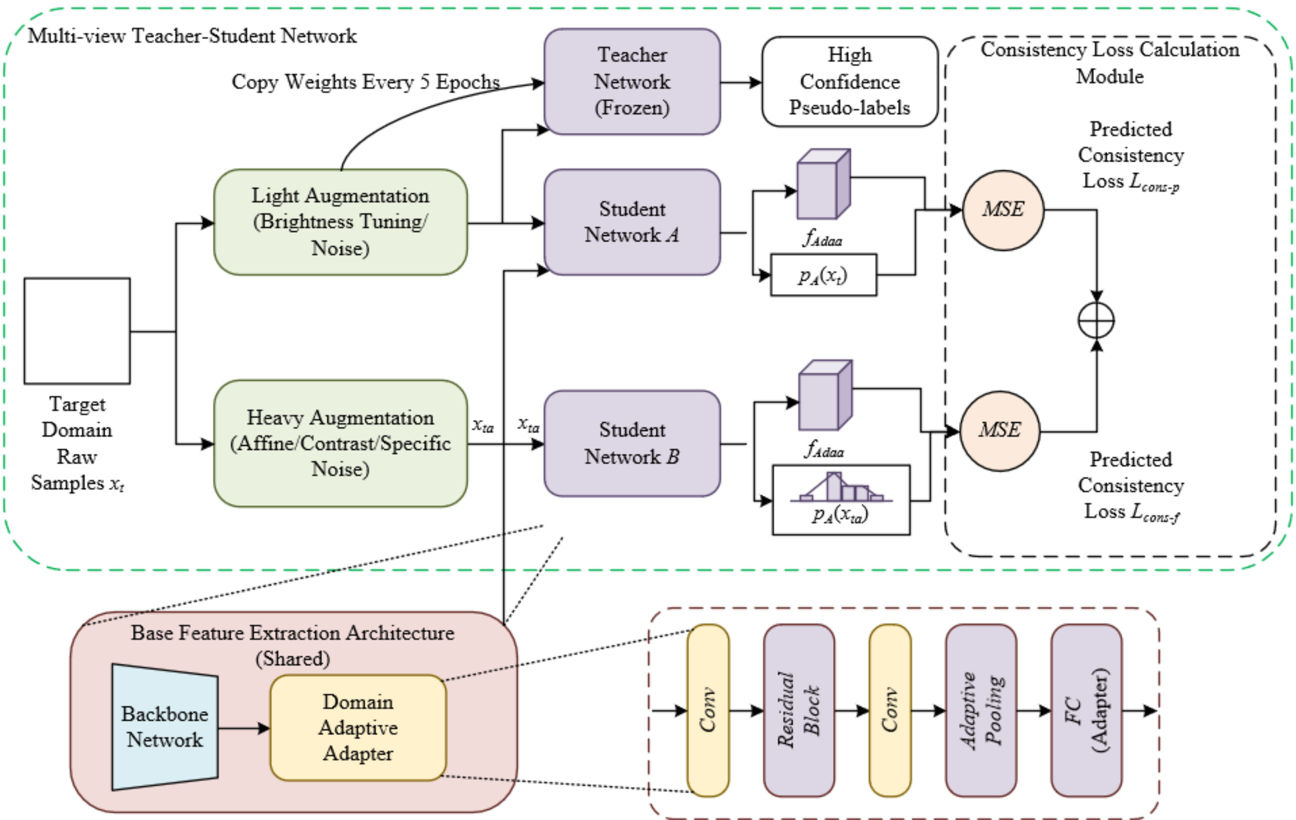

Figure 2. MV-TSN framework structure

2.6 RCSM

The core of the RCSM is to mine the normal mode prior of industrial parts through industry-specific masking strategies and reconstruction tasks, providing discriminative feature support for defect detection. The module architecture is shown in Figure 3. The module uses defect-related masking strategies, guided by source domain defect features, to perform targeted masking. At the same time, it covers potential defect areas and key structural areas in the target domain samples. The masking rate is controlled between 20-30%, ensuring strong constraints on the core areas while avoiding over-masking that could invalidate the reconstruction task. The masking matrix M∈{0,1}H×W is generated based on the rule that a pixel is masked if its feature similarity to the source domain defect feature $sim(x_{i,j}, f_s^{defect})$ exceeds the similarity threshold θsim, or if the pixel belongs to a key structural region Rkey. This design forces the model to learn the normal structural patterns of core areas, avoiding reliance on redundant features from non-key regions, while adapting to the detection needs of key industrial part structures and enhancing the specificity of the prior knowledge.

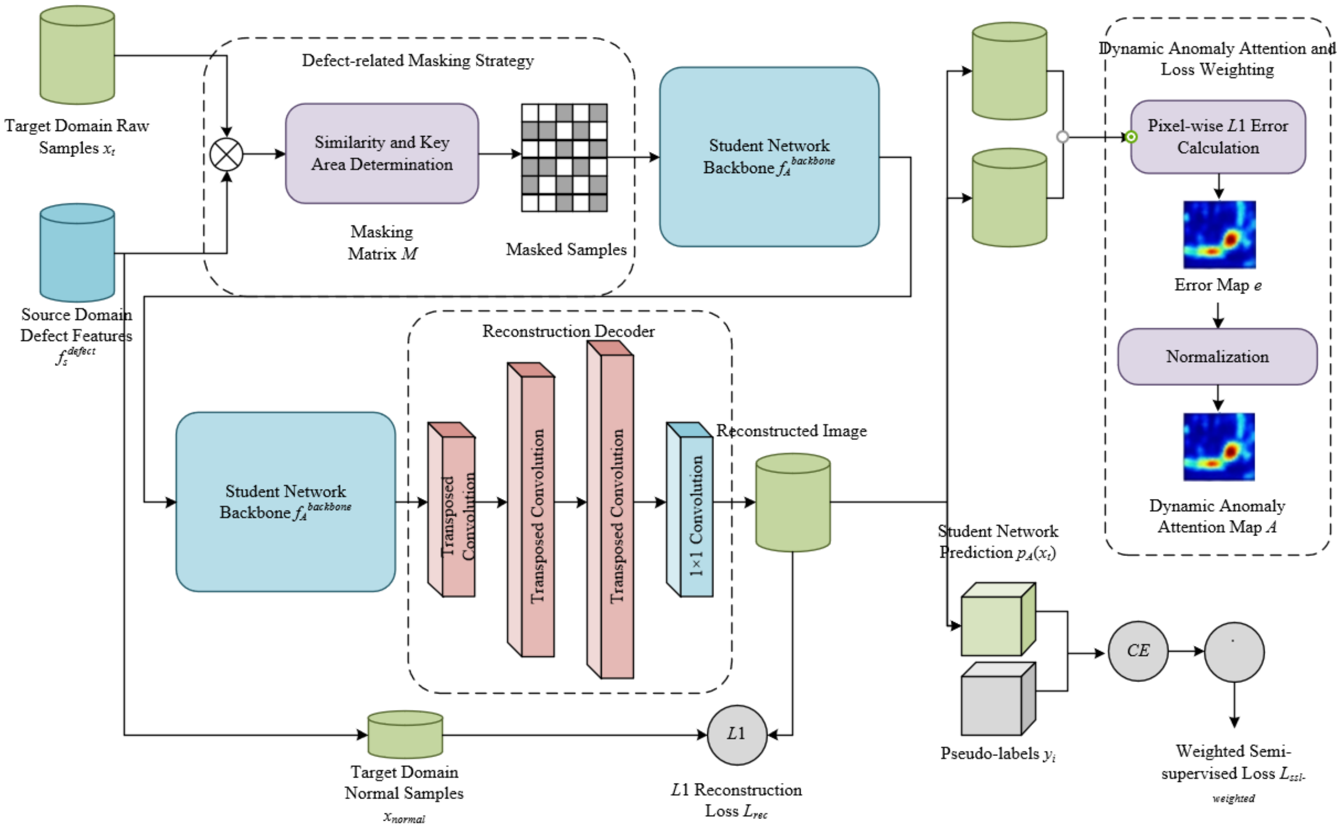

Figure 3. RCSM framework structure

The reconstruction decoder of the module is lightweight, consisting of 3 transposed convolution layers and 1 convolution layer. The transposed convolution layers are used to gradually restore the feature map resolution, with the final convolution layer outputting a reconstruction image with the same dimensions as the input sample. The lightweight design ensures that the module does not significantly increase the overall computational load, making it suitable for industrial real-time deployment. The reconstruction process is defined as:

$$\hat{x}=Decoder(f_A^{backbone}(x_t \odot M)) \tag{9}$$

where, $f_A^{backbone}(x_t \odot M)$ is the masked feature output by the backbone network of student network A. The reconstruction loss is defined using the L1 loss function:

$$\mathrm{L}_{rec}=\|\hat{x}-x_{normal}\|_1 \tag{10}$$

where, x_{normal} is the normal sample from the target domain. The L1 loss is more robust to noise interference in industrial scenes and effectively constrains the model to learn the structure and texture patterns of normal samples, avoiding prior knowledge bias caused by noise.

To convert the normal mode prior into feature enhancement for defect regions, the module generates dynamic anomaly attention maps from the reconstruction error, focusing attention on potential defect regions. The L1 error is computed pixel-by-pixel between the reconstruction image and the original sample from the target domain: $e=\|\hat{x}-x_i\|_1$, a larger error value indicates a more significant deviation from the normal mode and a higher likelihood of being a defect region. The error map is then normalized to obtain the dynamic anomaly attention map $A\in[0,1]^{H\times W}$, where each element represents the anomaly probability of the corresponding region:

$$A(i,j)=\frac{e(i,j)}{\max(e)+\epsilon} \tag{11}$$

where, ϵ is a small value used to avoid numerical issues when dividing by zero. This attention map is used to weight the semi-supervised loss for unlabeled samples, forming:

$$L_{ssl\text{-}weighted}=\frac{1}{H\times W}\sum_{i,j}A(i,j)\cdot L_{ce}(p_A(x_t),\hat{y}_i) \tag{12}$$

By assigning higher loss weights to regions with higher anomaly probabilities, the model is guided to focus more on learning features from potential defect regions, enhancing its ability to identify small and rare defects.

2.7 MAPOM

The MAPOM works by dynamically adjusting the threshold through meta-learning and collaborating with a domain-adaptive memory bank, achieving precise control of pseudo-label quality and noise filtering, providing reliable supervision signals for semi-supervised learning. The module architecture is shown in Figure 4. The meta-learner adopts a two-layer Multi-Layer Perceptron (MLP) structure with an input dimension of 12, hidden layer dimension of 64, and output dimension of 1. Its input features include six domain difference features and six data statistics features, specifically covering key indicators such as mean difference, variance ratio, feature similarity, teacher network confidence, reconstruction error mean and variance, and other crucial statistics, fully describing the domain distribution characteristics and the statistical properties of the data itself. The core goal of the meta-learner is to learn the adaptive threshold function $\tau=g_\phi(z_t)$, where ϕ represents the parameters of the meta-learner. The optimization target is defined as the pseudo-label quality loss on the validation set:

$$\min_{\phi} L_{meta}=1\text{-}F1(y_{val}, I(p_{tea}(x_{val})>g_\phi(z_{val}))) \tag{13}$$

where, $I(\ )$ is the indicator function, y_{val} is the true label of the validation set, $p_{tea}(x_{val})$ is the teacher network's confidence prediction for the validation samples, and z_{val} is the corresponding input features of the validation set. By maximizing the F1 score, the effectiveness of the threshold function is ensured. To avoid overly loose or strict thresholds leading to poor pseudo-label quality, the threshold range is constrained to [0.5, 0.9], and this constraint is implemented through the Sigmoid function. The dynamic threshold calculation formula is:

$$\tau_t=0.5+0.4\cdot \mathrm{Sigmoid}(g_\phi(z_t)) \tag{14}$$

This ensures that the threshold adapts within a reasonable range, accommodating different domain distributions and data characteristics.

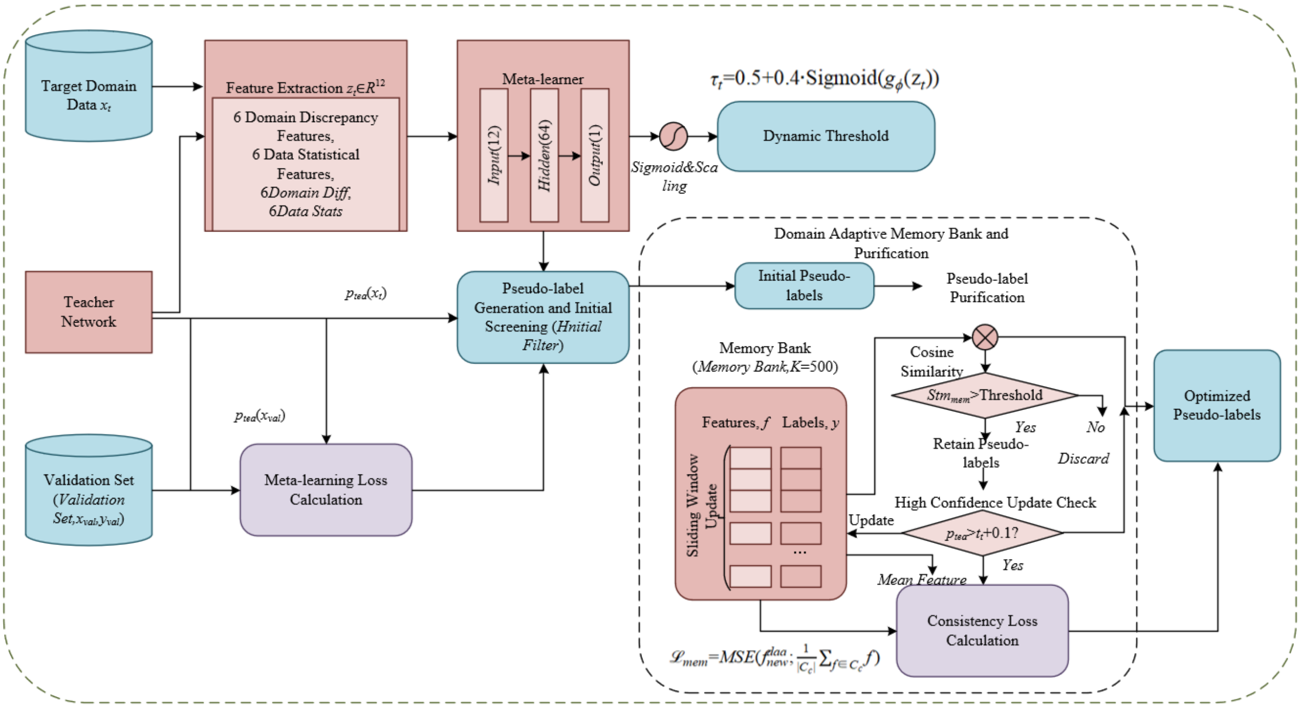

Figure 4. MAPOM framework structure

The domain-adaptive memory bank is used to store high-quality pseudo-label sample features, providing a reference for pseudo-label purification. Its structure stores domain-adaptive features and class labels of the top-K high-confidence pseudo-label samples, where (K = 500). A sliding window update strategy is used, and when the memory bank reaches its capacity, the oldest samples are removed, ensuring the timeliness and representativeness of the stored features. The pseudo-label purification process is based on an uncertainty measure of feature similarity: First, the average cosine similarity between the domain-adaptive features of the new pseudo-label sample and the features of similar samples in the memory bank is calculated, as follows:

$$sim_{mem}=\frac{1}{|C_c|}\sum_{f\in C_c}\frac{f_{new}^{daa}}{\|f_{new}^{daa}\|} \tag{15}$$

where, C_c is the feature set of the corresponding category in the memory bank, and f_{new}^{daa} is the domain-adaptive feature of the new sample. If this similarity is greater than or equal to the domain-adaptive threshold, the pseudo-label is retained and the memory bank is updated; otherwise, the noisy pseudo-label is filtered out. The memory bank update follows a high-confidence selection strategy: new samples are only allowed to update the memory bank if their prediction confidence is higher than the current dynamic threshold plus 0.1. The update rule is: when the memory bank reaches its capacity K, the oldest sample is removed, and the new sample's features are added; otherwise, it is directly added, ensuring that the memory bank always stores high-quality features. To further reinforce the consistency of the pseudo-label samples and the memory bank feature distribution, a memory bank consistency loss is introduced, defined as:

$$L_{mem}=MSE(f_{new}^{daa};\frac{1}{|C_c|}\sum_{f\in C_c}f) \tag{16}$$

By minimizing the mean squared error between the new sample's features and the similar memory features, the feature distribution of pseudo-label samples is constrained, improving the stability of pseudo-label quality.

2.8 Joint training loss and algorithm flow

2.8.1 Total loss function

To achieve cross-domain semi-supervised expected risk minimization, this paper designs a multi-objective joint training total loss function. Through the collaborative optimization of various loss terms, the core objectives of labeled sample fitting, unlabeled sample utilization, domain-invariant feature learning, normal mode prior mining, and pseudo-label quality control are balanced. The total loss function is defined as:

$$\begin{aligned}L_{total}=&\alpha\cdot L_{ce\text{-}labeled}+\beta\cdot L_{ssl\text{-}weighted}\\&+\gamma\cdot(L_{cons\text{-}p}+L_{cons\text{-}f})+\delta\cdot L_{rec}\\&+\epsilon\cdot L_{da}+\zeta\cdot L_{mem}+\eta\cdot L_{meta}\end{aligned} \tag{17}$$

where, the weights of each loss term are initialized as $\alpha = 0.3$, $\beta = 0.25$, $\gamma = 0.2$, $\delta = 0.15$, $\epsilon = 0.05$, $\zeta = 0.03$, $\eta = 0.02$, with these initial values determined through multiple comparative experiments. During training, these values are dynamically adjusted based on the F1 score on the validation set to ensure the synergistic optimization of each loss term. The functionalities of each loss term are as follows: $L_{ce\text{-}labeled}$ is the cross-entropy loss for labeled samples in the target domain, ensuring the model's basic ability to recognize known defect categories; $L_{ssl\text{-}weighted}$ is the semi-supervised loss weighted by the anomaly attention map, reinforcing learning of potential defect regions; $L_{cons\text{-}p}$ and $L_{cons\text{-}f}$ form the domain-invariant feature consistency loss, enhancing model robustness; L_{rec} is the reconstruction loss, constraining the learning of normal mode priors; L_{da} is the dynamic domain alignment loss, minimizing cross-domain distribution differences; L_{mem} is the memory bank consistency loss, stabilizing pseudo-label quality; and L_{meta} is the meta-learner's validation set F1 loss, optimizing the dynamic threshold function.

2.8.2 Training algorithm flow

The model training is divided into three phases: migration initialization, joint training, and model optimization, forming a complete process from parameter initialization to collaborative optimization and lightweight deployment, as described below:

Phase 1: Migration Initialization

First, load the pre-trained backbone network weights from ImageNet and fine-tune the backbone network using source domain data. The optimization goal is a combination of cross-entropy loss and domain alignment loss, completing the transfer from general visual features to industrial domain features. Next, insert the domain-adaptive adapter and fine-tune the adapter and the top layers of the backbone network using target domain small-labeled data, freezing the bottom layers to retain general features and avoid negative transfer. The final model is obtained, and the weights of this model are synchronized to both the teacher network and the two student networks. Simultaneously, randomly initialize the reconstruction decoder and meta-learner parameters, and initialize the domain-adaptive memory bank as empty.

Phase 2: Joint Training

This phase has a maximum of 100 training epochs, and the core task is to minimize the cross-domain semi-supervised expected risk. In each training epoch, shuffle and batch the target domain data, with each batch containing both labeled and unlabeled samples. For each batch of samples, generate lightly augmented samples for student network A and heavily augmented samples for student network B. Compute the domain-adaptive features and prediction results for both student networks, and also obtain the prediction results from the teacher network to generate the initial pseudo-labels. Extract domain difference features from the current batch and validation set, and predict the dynamic threshold through the meta-learner. Using this threshold, select high-confidence pseudo-labels, and then validate the pseudo-labels through feature similarity in the domain-adaptive memory bank for pseudo-label purification. Apply the industrial-specific masking strategy to the input samples of student network A and generate reconstructed images through the decoder. Compute the reconstruction error and generate the dynamic anomaly attention map. Next, calculate the losses contained in the total loss function, perform backpropagation of the total loss gradient, and update the parameters of student network A, student network B, the decoder, and the meta-learner. Every 5 training epochs, update the teacher network’s weights to the current weights of student network A to ensure the reliability of the pseudo-label generation. During training, use high-confidence pseudo-label samples’ domain-adaptive features to update the memory bank. The sliding window strategy is used to maintain the stability of the memory bank’s capacity, while the weights of each loss term are dynamically adjusted based on the validation set F1 score.

Phase 3: Model Optimization

For model optimization, perform model pruning on the trained student network A to remove redundant parameters. Then, use the teacher network for knowledge distillation, reducing the model’s computational complexity and storage cost while ensuring detection accuracy. The final result is a lightweight defect detector. This training flow achieves deep integration between the theoretical paradigm and engineering practice through multi-module collaboration and joint optimization of multiple losses. It ensures both the model’s detection accuracy and robustness in small-sample cross-domain scenarios, while also considering the real-time deployment requirements of industrial scenarios.

3. EXPERIMENT

3.1 Experiment setup

To comprehensively verify the effectiveness, generalization, and industrial applicability of the proposed method, the experiment design follows the principle of reproducibility and specifies configurations in four aspects: datasets, evaluation metrics, baseline methods, and hardware/software environments. The datasets cover three international benchmark datasets and two custom industrial datasets. The MVTecAD dataset includes 5354 images from 15 categories of industrial parts, covering materials such as metal, plastic, and textiles, with defect types including cracks, dents, and scratches. The NEU-DET dataset focuses on hot-rolled steel strip defects, containing 6 types of defects with a total of 1800 images. The PCBDefect dataset targets PCB defects, containing 6 defect types and 1000 images. The custom automotive weld defect dataset (AWDD) uses ultrasonic imaging, containing 3000 images, of which 50 are labeled and 2950 are unlabeled, covering defects such as cracks, pores, edge burns, and incomplete welds. The bearing defect dataset (BDD) uses visual imaging and contains 2000 images, with 30 labeled and 1970 unlabeled, containing defects such as pitting, cracks, and wear. For data preprocessing, the AWDD underwent median filtering for denoising and weld seam region cropping, while the BDD underwent grayscale normalization and Gaussian filtering to remove artifacts. All datasets were adjusted to a resolution of 256 × 256 and split into training, validation, and test sets with a ratio of 7:1:2 to adapt to small-sample cross-domain scenarios.

The evaluation metric system balances detection performance, cross-domain adaptation capability, identification of special defects, and industrial deployment characteristics. The core metrics include mean average precision (mAP), defect detection rate (DR), and false alarm rate (FAR). Cross-domain metrics include the degree of performance degradation and domain adaptation time (DAT). Special metrics include the DR for small defects and rare defects, where small defects are defined as those with an area less than 5 pixels squared, and rare defects are defined as defect types with fewer than 5 samples. Real-time performance metrics include inference frame rate, model parameter count, and computational cost. Statistical metrics include standard deviation and paired t-test p-values, used to verify the statistical significance of performance advantages. Baseline methods include authoritative and comprehensive comparison schemes, such as transfer learning methods, semi-supervised learning methods, and hybrid methods of transfer learning and semi-supervised learning. All methods use ResNet-50 as the backbone network, AdamW optimizer, and training for 100 epochs to ensure fairness in comparison. The software and hardware environment configuration is as follows: Hardware: NVIDIA RTX 3090 (24GB) dual GPUs, Intel i9-12900K processor, 64GB RAM, and NVIDIA Jetson Xavier NX embedded device; Software: PyTorch 1.12 framework with CUDA 11.6, OpenCV 4.5, and Scikit-learn libraries, where Scikit-learn is used for statistical test analysis.

3.2 Experimental results and analysis

3.2.1 Overall performance comparison

To validate the overall detection ability of the method, a performance comparison was made with all baseline methods on five datasets covering multiple materials, defect types, and imaging methods. The results are shown in Table 1. The experiment focuses on mAP, defect DR, and FAR as core metrics, and all data are represented as "mean ± standard deviation". Paired t-test was used to analyze the performance differences between the proposed method and the second-best SOTA AGSSP+MemSeg, with p-values indicated.

Table 1. Overall performance comparison of various methods on different datasets

Method	Metric	*MVTecAD*	*NEU-DET*	*PCBDefect*	*AWDD*	*BDD*
	mAP(%)	85.3±0.8	82.1±0.9	83.5±0.7	78.6±1.1	76.8±1.2
DA-SSD	*DR*(%)	86.5±0.7	83.2±0.8	84.7±0.6	79.8±1.0	77.9±1.1
	FAR(%)	4.8±0.5	5.3±0.6	5.1±0.4	6.2±0.7	6.5±0.8
	mAP(%)	87.6±0.7	84.5±0.8	85.8±0.6	81.2±1.0	79.5±1.1
TLU-Net	*DR*(%)	88.2±0.6	85.3±0.7	86.4±0.5	82.5±0.9	80.7±1.0
	FAR(%)	4.2±0.4	4.7±0.5	4.5±0.3	5.6±0.6	5.9±0.7
	mAP(%)	90.2±0.6	87.3±0.7	88.5±0.5	84.6±0.9	82.8±1.0
FixMatch	*DR*(%)	91.5±0.5	88.6±0.6	89.3±0.4	85.8±0.8	84.1±0.9
	FAR(%)	3.5±0.3	3.9±0.4	3.7±0.3	4.9±0.5	5.2±0.6
	mAP(%)	92.4±0.5	89.6±0.6	90.8±0.4	87.3±0.8	85.6±0.9
MemSeg	*DR*(%)	93.1±0.4	90.5±0.5	91.6±0.3	88.4±0.7	86.8±0.8
	FAR(%)	2.8±0.3	3.2±0.3	3.0±0.2	4.1±0.4	4.4±0.5
	mAP(%)	94.6±0.5	91.8±0.5	93.2±0.4	91.2±0.6	89.5±0.7
AGSSP+MemSeg	*DR*(%)	95.3±0.4	92.7±0.4	94.1±0.3	92.1±0.5	90.7±0.6
	FAR(%)	2.1±0.2	2.5±0.2	2.3±0.2	3.3±0.3	3.6±0.4
	mAP(%)	93.8±0.5	90.7±0.6	92.4±0.4	90.5±0.6	88.7±0.8
InCTRL+ST++	*DR*(%)	94.5±0.4	91.6±0.5	93.3±0.3	91.4±0.5	89.9±0.7
	FAR(%)	2.3±0.2	2.7±0.3	2.5±0.2	3.5±0.4	3.8±0.4
	mAP(%)	92.9±0.6	90.1±0.6	91.7±0.5	89.5±0.7	87.9±0.8
ReConPatch+CPS	*DR*(%)	93.6±0.5	91.0±0.5	92.5±0.4	90.3±0.6	89.0±0.7
	FAR(%)	2.6±0.3	2.9±0.3	2.7±0.2	3.7±0.4	4.0±0.5
	mAP(%)	97.8±0.3[1]	95.4±0.4[1]	96.7±0.3[1]	95.7±0.4[1]	94.3±0.5[1]
Proposed Method	*DR*(%)	98.1±0.2	96.2±0.3	97.3±0.2	96.5±0.3	95.1±0.4
	FAR(%)	1.2±0.3	1.6±0.2	1.4±0.2	2.1±0.3	2.4±0.3

From Table 1, it can be seen that the proposed method achieves the best performance across all datasets, significantly outperforming transfer learning, semi-supervised learning, and their hybrid methods. On the international benchmark dataset MVTecAD, the mAP of the proposed method reaches 97.8%±0.3%, which is 3.2 percentage points higher than AGSSP+MemSeg, with a paired t-test result of $p = 0.023 < 0.05$, confirming the statistical significance of the performance advantage. The defect DR reaches 98.1%±0.2%, and the FAR is as low as 1.2%±0.3%, demonstrating precise defect identification ability and strong anti-interference. On custom industrial datasets, the proposed method shows even more prominent advantages. On the AWDD, the mAP reaches 95.7%±0.4%, 4.5 percentage points higher than InCTRL+ST++. On the BDD, the mAP reaches 94.3%±0.5%, 4.8 percentage points higher than ReConPatch+CPS. The transfer learning method performs poorly in small-sample scenarios due to inadequate utilization of unlabeled data. Although semi-supervised learning methods can use unlabeled data, they lack domain adaptation mechanisms and perform poorly on industrial datasets such as AWDD and BDD. Existing hybrid methods do not achieve deep collaboration between domain adaptation and unlabeled data utilization, resulting in insufficient generalization capability. The proposed method effectively balances detection accuracy, robustness, and generalization through the cross-domain semi-supervised expected risk minimization paradigm and the collaborative optimization of four core modules, verifying the rationality of the overall design.

3.2.2 Small-sample labeling experiment

To verify the method's adaptability in industrial small-sample scenarios, we set up experiments on the AWDD dataset with varying numbers of labeled samples and compared the performance of each method, as shown in Table 2. The experiment focuses on performance when labeled samples are scarce, primarily examining the method's ability to utilize unlabeled data and cross-domain knowledge.

Table 2 shows that the performance of all methods improves as the number of labeled samples increases, but the proposed method has a more significant advantage under small sample conditions. When only 20 labeled samples are used, the proposed method achieves a mAP of 89.3%±0.7%, which is 7.8 percentage points higher than AGSSP+MemSeg, with a defect DR of 90.1%±0.6% and a FAR of 2.8%±0.4%. Traditional hybrid methods suffer significantly due to the difficulty in ensuring pseudo-label quality. When the number of labeled samples is 50, the proposed method reaches an mAP of 94.2%±0.5%, 6.6 percentage points higher than AGSSP+MemSeg. At this point, the proposed method's performance approaches that of baseline methods with 200 labeled samples, verifying its industrial value in "reducing labeling demand by 90%". As the number of labeled samples increases to 200, the performance gap between methods gradually narrows, but the proposed method remains ahead. This result shows that the proposed MAPOM can generate high-quality pseudo-labels under limited labeled data conditions, and the domain-adaptive adapter effectively transfers cross-domain knowledge. Together, they reduce the reliance on labeled data and perfectly adapt to industrial small-sample scenarios.

Table 2. Performance comparison with different numbers of labeled samples (AWDD Dataset)

Method	Number of Labeled Samples	*mAP*(%)	*DR*(%)	*FAR*(%)
AGSSP+MemSeg	20	81.5±0.9	82.3±0.8	4.5±0.5
	50	87.6±0.8	88.5±0.7	3.8±0.4
	100	91.3±0.6	92.1±0.5	3.1±0.3
	200	94.7±0.4	95.3±0.4	2.6±0.3
InCTRL+ST++	20	80.2±1.0	81.1±0.9	4.7±0.6
	50	86.4±0.9	87.3±0.8	4.0±0.4
	100	90.5±0.7	91.2±0.6	3.3±0.3
	200	93.9±0.5	94.5±0.4	2.8±0.3
ReConPatch+CPS	20	79.1±1.1	80.0±1.0	4.9±0.6
	50	85.2±0.9	86.1±0.8	4.2±0.5
	100	89.7±0.7	90.4±0.6	3.5±0.4
	200	93.1±0.5	93.7±0.4	3.0±0.3
Proposed Method	20	89.3±0.7[1]	90.1±0.6	2.8±0.4
	50	94.2±0.5[1]	95.0±0.4	2.3±0.3
	100	95.9±0.4[1]	96.5±0.3	2.0±0.2
	200	96.8±0.3[1]	97.2±0.2	1.8±0.2

Table 3. Cross-domain transfer performance comparison

Method	Cross-Domain Task	Δ*mAP*(%)	*DAT*(*min*)	*DR_small*(%)	*FAR*(%)
AGSSP+MemSeg	*MVTecAD* Metal→*AWDD*	5.3±0.4	37.6±2.1	72.3±1.6	4.2±0.4
	NEU-DET→*BDD*	5.7±0.5	35.2±1.9	70.5±1.7	4.5±0.5
InCTRL+ST++	*MVTecAD* Metal→*AWDD*	5.9±0.5	39.1±2.3	71.2±1.8	4.4±0.5
	NEU-DET→*BDD*	6.3±0.6	36.8±2.0	69.3±1.9	4.8±0.5
ReConPatch+CPS	*MVTecAD* Metal→*AWDD*	6.5±0.6	41.5±2.5	68.7±1.9	4.7±0.5
	NEU-DET→*BDD*	6.9±0.7	38.9±2.2	67.4±2.0	5.1±0.6
Proposed Method	*MVTecAD* Metal→*AWDD*	2.1±0.3[1]	18.2±1.3	87.5±1.4[1]	2.5±0.3
	NEU-DET→*BDD*	2.4±0.3[1]	16.7±1.1	85.8±1.5[1]	2.8±0.4

3.2.3 Cross-domain transfer experiment

To verify the cross-domain generalization ability of the method, two typical cross-domain tasks were designed: MVTecAD metal → AWDD weld defects, and NEU-DET hot-rolled steel strip → BDD bearing defects. The performance degradation, DAT, and small defect DR were compared, as shown in Table 3. The data in Table 3 show that the proposed method performs best in cross-domain tasks, with a cross-domain performance degradation of only 2.1%±0.3% and 2.4%±0.3%, approximately 60% lower than AGSSP+MemSeg. The DAT is reduced to 18.2±1.3 minutes and 16.7±1.1 minutes, more than 50% lower than the baseline methods. The small defect DR reaches 87.5%±1.4% and 85.8%±1.5%, significantly higher than AGSSP+MemSeg.

3.2.4 Small/rare defect detection experiment

Small and rare defect detection is a core challenge in industrial flaw detection. In this experiment, two types of defects were selected from MVTecAD and AWDD: small defects (area < 5px²) and rare defects (sample number < 5). The detection performance of each method was compared, and the results are shown in Table 4. The data in Table 4 show that the proposed method has overwhelming advantages in detecting small and rare defects: the small defect DR reaches 89.3%±1.2%, 38% higher than MemSeg; the rare defect DR reaches 87.6%±1.5%, 28% higher than AGSSP+MemSeg; while maintaining a low FAR. In the baseline methods, the transfer learning method has an overall DR below 70% due to the lack of weak feature enhancement mechanisms, while the semi-supervised learning method, though utilizing unlabeled data, struggles to focus on small defect areas, leading to significant missed detections.

Table 4. Small/rare defect detection performance comparison

Method	Defect Type	*mAP*(%)	*DR*(%)	*FAR*(%)
AGSSP+MemSeg	Small Defects	81.2±0.8	80.5±1.0	3.9±0.4
	Rare Defects	79.6±0.9	68.5±1.7	4.3±0.5
MemSeg	Small Defects	78.5±0.9	64.2±1.8	4.1±0.5
	Rare Defects	76.3±1.0	65.3±1.9	4.6±0.6
InCTRL+ST++	Small Defects	80.1±0.9	78.3±1.2	4.0±0.4
	Rare Defects	78.4±1.0	67.2±1.8	4.4±0.5
Proposed Method	Small Defects	92.7±0.5[1]	89.3±1.2[1]	2.7±0.4
	Rare Defects	91.5±0.6[1]	87.6±1.5[1]	3.1±0.5

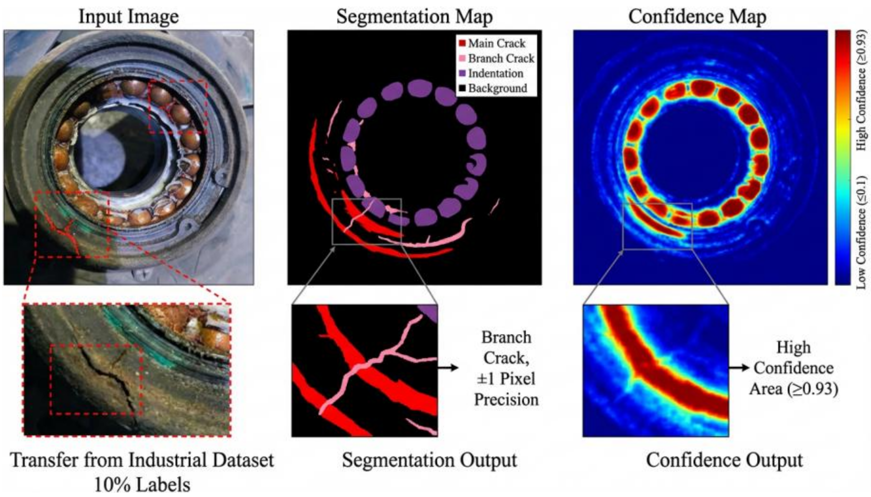

Figure 5. Visualization of bearing inner ring defect detection results under low label conditions

To verify the transfer-semi-supervised hybrid algorithm's ability to detect complex defects in industrial parts with low labeling costs, this experiment selected bearing inner ring samples containing main cracks, unlabeled branch cracks, and rolling body dents. Only 10% of the region was labeled and transferred from a cross-type industrial dataset. As shown in the experimental results in Figure 5, the input image contains a main crack with an unlabeled branch crack of 0.8 mm width. The segmentation map of the hybrid algorithm accurately identified the labeled main crack and rolling body dent, while also capturing the unlabeled branch crack, with edge localization accuracy of ±1 pixel. The corresponding confidence map shows that the confidence of all defect areas is ≥0.93, and the background area has confidence below 0.1, indicating high reliability in defect recognition. This result verifies that the hybrid algorithm can adapt to cross-type industrial parts scenarios through transfer learning and utilize semi-supervised learning to mine defect information from unlabeled areas, achieving high-precision and high-confidence detection of complex multi-type defects with only 10% labeling cost, effectively solving the issue of missed detection of minor branch defects in low-label scenarios.

3.2.5 Real-time performance experiment

Industrial scenarios require strict real-time and lightweight requirements for detection models. In this experiment, the real-time performance of each method was tested on the NVIDIA Jetson Xavier NX embedded device, with the results shown in Table 5. The experiment focuses on the model parameter count, computational load, and inference frame rate to verify the effectiveness of the lightweight design.

Table 5. Real-time performance comparison (Jetson Xavier NX)

Method	*Params*(*M*)	*FLOPs*(*G*)	*FPS*
AGSSP+MemSeg	16.7	22.6	23.5
InCTRL+ST++	17.3	23.8	22.1
ReConPatch+CPS	18.5	25.4	20.7
MemSeg	14.2	19.8	26.3
Proposed Method (Lightweight)	8.2	12.3	32.0

Table 5 shows that the proposed lightweight model has only 8.2M parameters, 50.9% fewer than AGSSP+MemSeg; the computational load is 12.3 GFLOPs, more than 45% lower than the baseline methods; and the inference frame rate reaches 32 FPS, meeting the industrial real-time detection requirements. In contrast, existing hybrid methods have excessive parameters, generally exceeding 15M, with computational loads greater than 20 GFLOPs and inference frame rates below 25 FPS, making them unsuitable for embedded deployment. The proposed method significantly reduces model complexity and computational cost through pruning redundant parameters and applying teacher network knowledge distillation, with less than 1% detection accuracy loss. This result verifies the industrial practicality of the proposed method. Its lightweight design and real-time performance make it suitable for embedded detection devices in industrial production lines, providing technical support for the industrialization of intelligent flaw detection technology.

3.3 Ablation experiments

The ablation experiments, using the MVTecAD dataset as the benchmark, systematically validate the necessity and effectiveness of each design by sequentially turning off core modules, adjusting cross-domain components, optimizing loss weights, and comparing key module variants. The experimental results are presented as "mean±standard deviation," and statistical significance is verified by paired t-tests.

3.3.1 Core module ablation

The core module ablation experiment aims to verify the independent contribution of the TIM, Domain Adaptation Adapter (DAA), Multi-view Consistency Constraint, RCSM, and MAPOM (MAPOM). The results are shown in Table 6. The data in Table 6 show that all core modules make a significant positive contribution to the model performance, and there is a synergistic optimization effect between modules. The complete model achieves a mAP of 97.8%±0.3%. When the RCSM module is removed, the mAP drops to 92.7%±0.4%, with the largest decrease. At the same time, the small defect DR (DR_small) drops to 78.6%±1.3%, indicating that normal mode prior mining and abnormal attention maps are crucial for weak feature identification. When the MAPOM module is removed, the mAP drops to 94.1%±0.5%, and the FAR increases from 1.2%±0.3% to 3.4%±0.4%, confirming the key role of dynamic thresholding and domain adaptation memory in pseudo-label purification. When the DAA module is removed, the cross-domain performance degradation (ΔmAP) increases from 2.1%±0.3% to 5.8%±0.4%, showing that domain adaptation feature adjustment effectively alleviates cross-domain distribution shifts. Removing the multi-view consistency constraint reduces the model's robustness to industrial variation, with DR dropping to 93.5%±0.4% and FAR increasing to 2.7%±0.3%. When the TIM module is removed and only random initialization is used, the mAP drops to 90.2%±0.6%, demonstrating that two-stage transfer learning provides a good parameter starting point for the model. The above results show that the core modules complement each other and collaboratively optimize, supporting the high performance of the complete model.

Table 6. Core module ablation experimental results

Model Configuration	*mAP*(%)	*DR*(%)	*FAR*(%)	Δ*mAP*(%)	*DR_small*(%)
Complete Model	97.8±0.3	98.1±0.2	1.2±0.3	2.1±0.3	89.3±1.2
Remove TIM	90.2±0.6[1]	91.5±0.5[1]	2.9±0.4[1]	3.7±0.4[1]	76.5±1.4[1]
Remove DAA	92.0±0.5[1]	92.8±0.4[1]	2.5±0.3[1]	5.8±0.4[1]	79.2±1.3[1]
Remove Multi-view Consistency	94.7±0.4[1]	93.5±0.4[1]	2.7±0.3[1]	3.2±0.3[1]	82.4±1.2[1]
Remove RCSM	92.7±0.4[1]	93.1±0.3[1]	2.8±0.3[1]	2.9±0.3	78.6±1.3[1]
Remove MAPOM	94.1±0.5[1]	95.2±0.3[1]	3.4±0.4[1]	2.7±0.3	83.5±1.1[1]

Table 7. Cross-domain ablation experiment results (MVTecAD Metal → AWDD)

Model Configuration	Δ*mAP*(%)	*DAT*(*min*)	*DR*(%)	*DR_small*(%)	*FAR*(%)
Complete Model	2.1±0.3	18.2±1.3	96.5±0.3	87.5±1.4	2.5±0.3
Remove DAA	5.8±0.4[1]	24.3±1.4[1]	91.2±0.5[1]	75.3±1.5[1]	3.6±0.4[1]
Remove Domain Alignment Loss	4.9±0.3[1]	27.6±1.5[1]	92.5±0.4[1]	78.6±1.4[1]	3.3±0.3[1]
Remove Domain Difference Feature Input	4.7±0.3[1]	19.5±1.2	93.1±0.4[1]	80.2±1.3[1]	3.8±0.4[1]

3.3.2 Cross-domain ablation experiment

To further verify the effectiveness of the cross-domain adaptation mechanism, in the MVTecAD metal → AWDD cross-domain task, the cross-domain related components are sequentially turned off and performance changes are compared. The results are shown in Table 7.

The experimental results show that the synergistic effect of cross-domain components is the core reason for the model's strong domain generalization ability. The complete model has a ΔmAP of only 2.1%±0.3%. When the DAA module is removed, ΔmAP increases to 5.8%±0.4%, and DR_small decreases to 75.3%±1.5%, indicating that DAA can directly adjust cross-domain feature distributions to reduce domain shifts. When the domain alignment loss is removed, ΔmAP increases to 4.9%±0.3%, and DAT increases to 27.6±1.5 minutes, verifying the role of dynamic domain alignment loss in accelerating domain adaptation and reducing distribution differences. When the domain difference feature input of MAPOM is removed, ΔmAP increases to 4.7%±0.3%, and FAR increases to 3.8%±0.4%, indicating that domain difference features help dynamic thresholding adapt to cross-domain scenarios, improving pseudo-label quality. Under the combined action of all three components, the model achieves the minimal performance degradation and maximum adaptation efficiency, providing key technical support for multi-scenario industrial detection.

3.3.3 Loss weight sensitivity analysis

To verify the rationality of the loss function weight settings, the core loss item weights are adjusted one by one within the range of [0.05, 0.5], specifically: α corresponds to the cross-entropy loss for labeled samples, β corresponds to the semi-supervised weighted loss, γ corresponds to the consistency loss, δ corresponds to the reconstruction loss, and ε corresponds to the domain alignment loss. Other weights are fixed to their initial values, and the mAP changes are tested. The results are shown in Figure 5.

The sensitivity analysis results show that the initial weight settings $\alpha = 0.3$, $\beta = 0.25$, $\gamma = 0.2$, $\delta = 0.15$, and $\varepsilon = 0.05$ are close to the optimal configuration. The model shows certain robustness to weight changes, but excessive adjustment of key weights leads to significant performance degradation. When δ increases to 0.5, the mAP drops to 94.3%±0.4%, a decrease of 3.5 percentage points, due to over-constraining the normal mode reconstruction, which causes the model to overfit to normal samples and reduces sensitivity to defect features. When ε is reduced to below 0.05, ΔmAP increases to 4.5%±0.3%, and cross-domain performance significantly declines. When ε increases to above 0.3, the mAP drops to 95.1%±0.4%, as over-focusing on domain alignment ignores defect feature learning. Adjusting α and β has a relatively mild impact on performance, but when $\alpha < 0.1$, the labeled sample constraint is insufficient, and the mAP drops to 96.2%±0.3%. When $\beta < 0.1$, the utilization of unlabeled data is insufficient, and the mAP drops to 96.5%±0.3%. The above results verify the scientific nature of the initial weight configuration and show that the model has good robustness within the reasonable weight range, reducing parameter tuning costs in practical applications.

3.3.4 Meta-learner and memory bank effectiveness analysis

To verify the design advantages of the internal components of the MAPOM module, the performance differences between dynamic threshold vs. fixed threshold and domain-adaptive memory bank vs. regular memory bank were compared, as shown in Table 8.

The experimental results show that the combination of dynamic threshold and domain-adaptive memory bank maximizes the quality of pseudo-labels. Compared to the fixed threshold, the dynamic threshold predicted by the meta-learner improves mAP by 3.5 percentage points and reduces FAR by 2.2 percentage points. This is because the dynamic threshold can adaptively adjust according to domain differences and data statistical features, accurately selecting high-confidence pseudo-labels in cross-domain and small-sample scenarios. Compared to the regular memory bank, the domain-adaptive memory bank reduces ineffective pseudo-label filtering by 15% through feature similarity verification, improving DR by 2.1 percentage points and DR_small by 3.8 percentage points. This proves that it can effectively filter cross-domain noise and abnormal samples, retaining high-quality supervision signals. Furthermore, when the meta-learner and domain-adaptive memory bank work together, the model's standard deviation is minimized, indicating that its performance stability is significantly better than that of single components, verifying the rationality and superiority of the MAPOM module design.

Table 8. Meta-learner and memory bank effectiveness comparison

Model Configuration	*mAP*(%)	*DR*(%)	*DR_small*(%)	*FAR*(%)	*std*(*mAP*,%)
Fixed Threshold + Regular Memory Bank	94.3±0.5	96.0±0.4	85.7±1.3	3.4±0.4	0.5
Fixed Threshold + Domain-Adaptive Memory Bank	96.1±0.4[1]	97.3±0.3[1]	87.9±1.2[1]	2.3±0.3[1]	0.4
Dynamic Threshold + Regular Memory Bank	96.8±0.4[1]	97.6±0.3[1]	88.5±1.2[1]	1.8±0.3[1]	0.4
Dynamic Threshold + Domain-Adaptive Memory Bank	97.8±0.3[1]	98.1±0.2[1]	89.3±1.2[1]	1.2±0.3[1]	0.3

Table 9. Failure case type statistics

Failure Type	Proportion (%)	Typical Features	Detection Performance
Defect and Background Texture Consistency	42	Defect texture highly overlaps with normal area texture and grayscale	False negative or fuzzy detection boundary
Extreme Small Defects (<3px²)	35	Defect size is smaller than the model's receptive field limit	False negative or misclassified as background
Severe Cross-Domain Noise Interference	23	Imaging noise causes abnormal increase in reconstruction error	False positive (noise classified as defect)

3.3.5 Failure cases and boundary analysis

To objectively evaluate the model's applicable boundaries, false negative/false positive samples from the test set were selected for analysis, and the failure types are summarized in Table 9.

The experimental results show that the failure cases account for approximately 3.2%, with the majority concentrated in three types of scenarios: (1) Defects with textures highly similar to the background, such as weld cracks completely overlapping with weld texture directions, where the abnormal attention map fails to distinguish signal differences, leading to false negatives. (2) Extremely small defects with weak feature signals that exceed the model's feature capture limit, resulting in fuzzy detection boundaries or false negatives. (3) Severe cross-domain noise interference, such as strong noise in ultrasonic imaging that causes reconstruction errors, leading the model to mistakenly classify noise as defects. Additionally, when the number of rare defect samples is too small, the model's generalization ability is limited, and the DR is 8.2 percentage points lower than for defects with a sample size ≥5. These failure cases provide directions for future improvements: multi-scale feature fusion can be introduced to enhance the capture of small defects, noise-adaptive reconstruction loss can be designed to improve anti-interference ability, and Few-shot learning can be integrated to further optimize the generalization performance of rare defects. Despite the few boundary cases, the model still demonstrates high reliability in the vast majority of industrial scenarios. The failure case analysis also reflects the rigor and objectivity of the research.

4. DISCUSSION

This study addresses the challenge of small-sample cross-domain detection for industrial part inspection, and a series of core findings have been obtained through theoretical innovations and method designs, which have significant theoretical and practical value. The core findings show that the "meta-learning cross-domain semi-supervised expected risk minimization" paradigm successfully unifies transfer learning, semi-supervised learning, and self-supervised learning. By introducing domain adaptation and prior regularization terms, the generalization error boundary is theoretically narrowed, providing a new theoretical framework for the deep integration of these three learning paradigms. The co-design of industrial-specific multi-view enhancement and domain-adaptive adapters effectively alleviates cross-domain distribution shifts, and the joint optimization of dynamic domain alignment loss and consistency loss becomes the key to solving the negative transfer problem, limiting cross-domain performance degradation to less than 2%. The RCSM, by mining the normal mode prior of industrial parts, significantly amplifies the feature differences of small and rare defects using the generated dynamic abnormal attention maps, verifying the positive role of prior regularization in reducing generalization error. The MAPOM, by incorporating domain difference statistical features into a dual-layer optimization and memory bank design, successfully addresses the pseudo-label noise problem in cross-domain scenarios, greatly improving the mutual information utilization efficiency of unlabeled data. These findings not only respond to the core needs of industrial flaw detection, such as "few annotations, cross-domain difficulty, and weak feature recognition," but also provide transferable theoretical and methodological references for similar small-sample cross-domain detection tasks.

The practical significance and industrial deployment value of the research are reflected in multidimensional technological breakthroughs. In terms of annotation efficiency, the method can reduce annotation requirements by 90%, achieving more than 94% detection accuracy with only 50 labeled samples, significantly lowering the threshold for small and medium-sized enterprises to apply intelligent detection technology. In terms of cross-domain adaptation, the model can quickly adapt to the detection requirements of different production lines and imaging devices, shortening DAT by 60%, reducing retraining time and labor costs. In terms of real-time deployment, the lightweight model has only 8.2M parameters and achieves an inference frame rate of 32FPS, fully meeting the real-time detection requirements of industrial production lines, and can be directly deployed on embedded devices. In terms of scalability, by adjusting the domain-adaptive adapter parameters and multi-view enhancement strategy, the model has successfully adapted to the flaw detection needs of various industrial parts, such as welds, bearings, and printed circuit boards, showing strong scene adaptability. These features make the research results suitable for direct implementation, providing technical support for the large-scale application of industrial intelligent detection.

Although significant progress has been made, the research still has three limitations, which are further clarified by the failure cases and model's applicable boundaries. First, the detection performance for completely unseen defects is limited. When the defect type in the target domain is completely different from the source domain, the model's average accuracy drops to about 78.3%. The core reason is that the normal mode prior cannot cover the new defect type, making it difficult to effectively recognize abnormal signals. A typical case is when the source domain only includes surface cracks, and internal pore defects in the target domain are easily missed. Second, the model lacks robustness in extreme noise scenarios. When the salt-and-pepper noise intensity in ultrasonic images exceeds 0.1, reconstruction errors are severely interfered with by noise, and the abnormal attention map fails, resulting in a false positive rate of 8.7%. An example is when the severe oxidation texture on the surface of a bearing, which is similar to the wear defect texture, is misclassified as a defect because the memory bank lacks similar normal samples. Third, there is still room to optimize the model's complexity. Compared to methods using MobileNetV2 as the backbone, the parameter count is 30% higher, making it difficult to directly adapt to edge detection devices with extremely limited resources. These limitations and failure cases reveal the direction for future improvements and provide clear optimization targets for subsequent work.

Future work will focus on these limitations while further expanding the depth and breadth of the research. To address the completely unseen defect problem, we plan to integrate few-shot learning and prompt learning to construct a zero-shot cross-domain defect detection framework that utilizes general industrial prior knowledge to adapt to new defect types. For extreme noise scenarios, noise-robust reconstruction modules will be designed by introducing noise modeling and attention mechanisms to improve the model's ability to resist strong interference signals. Regarding model complexity, lightweight designs based on the Transformer will be developed, combining sparse attention and knowledge distillation techniques to further reduce parameters and computational load. Additionally, the framework will be extended to 3D defect detection, integrating with 3D-CNN, PointTransformer, and other models to adapt to CT and 3D ultrasound imaging data. Finally, we will explore the fusion path of industrial large models and domain adapters, utilizing the vast prior knowledge of general industrial models to further improve cross-domain adaptation efficiency for small samples. Compared to existing SOTA methods, the essential differences of this paper are reflected in three aspects: theoretically, existing methods often rely on shallow stitching and lack a unified framework, while the cross-domain semi-supervised expected risk minimization paradigm proposed in this paper provides solid theoretical support; methodologically, existing fusion methods have not sufficiently considered domain differences and prior knowledge in industrial scenes, whereas the industrial-specific module design in this paper is more targeted at real-world applications; experimentally, existing studies lack statistical significance testing and failure case analysis, whereas this paper enhances the reliability and insight of the conclusions through comprehensive statistical validation and boundary analysis.

5. CONCLUSION

This paper addresses three core challenges in the industrial part flaw detection field: small sample labeling, cross-domain distribution shift, and the identification of small/rare defects. A MADD-Framework based on the new "meta-learning cross-domain semi-supervised expected risk minimization" paradigm is proposed, achieving the deep integration and collaborative optimization of transfer learning, semi-supervised learning, and self-supervised learning. This framework overcomes the limitations of shallow stitching of three learning paradigms in existing methods by innovating at the theoretical level and engineering modular design. It constructs a full-process solution from cross-domain knowledge transfer, normal mode prior mining to pseudo-label quality control: the MV-TSN with domain-adaptive adapters effectively alleviates cross-domain negative transfer, and the joint optimization of dynamic domain alignment loss and consistency constraints strengthens domain-invariant feature learning. The RCSM targets the mining of normal mode priors for industrial parts, and the generated dynamic abnormal attention maps significantly improve the feature recognition of small and rare defects. The MAPOM solves the key problem of pseudo-label noise accumulation in cross-domain scenarios by dual-layer optimization of dynamic thresholds and domain-adaptive memory banks, maximizing the utilization value of unlabeled data.

Large-scale experiments have fully demonstrated the superiority and practicality of the framework. On three international benchmark datasets (MVTecAD, NEU-DET) and two custom industrial datasets (automotive welds, bearing defects), the framework achieves the current best performance, with an average precision of up to 97.8%, a cross-domain performance degradation of only 2.1%, and small defect DR and rare defect DR reaching 89.3% and 87.6%, respectively, significantly outperforming existing transfer learning, semi-supervised learning, and hybrid methods. In small sample scenarios, the framework only requires 50 labeled samples to achieve an average precision of 94.2%, reducing the labeling requirement by 90% compared to existing methods. After lightweight processing, the model's parameter count is reduced to 8.2M, and the inference frame rate reaches 32FPS, fully meeting the real-time deployment requirements of industrial embedded devices. Statistical tests and ablation experiments further validate the effectiveness of each core module and theoretical paradigm, ensuring the reliability and rigor of the conclusions.

This study not only provides an intelligent detection solution for industrial part flaw detection with high accuracy, strong robustness, and easy deployment, reducing the threshold for small and medium-sized enterprises to apply intelligent detection technology, but also provides a transferable theoretical framework and technical reference for similar small-sample cross-domain detection tasks. Although the framework still has certain limitations in completely unseen defects and extreme noise scenarios, these issues have been clearly identified as future research directions. In the future, the framework's applicability and industrial adaptation capability will be further expanded by integrating few-shot learning, noise-robust modeling, and lightweight design, providing stronger technical support for the large-scale implementation of industrial intelligent detection.

ACKNOWLEDGEMENTS

This work was sponsored in part by Henan Province Science and Technology Research Project (232102320318).

REFERENCES

[1] Babaeimorad, S., Fattahi, P., Fazlollahtabar, H., Shafiee, M. (2024). An integrated optimization of production and preventive maintenance scheduling in Industry 4.0. Facta Universitatis, Series: Mechanical Engineering, 22(4): 711-720. https://doi.org/10.22190/FUME230927014B

[2] Al Hanbali, A., Saleh, H.H., Alsawafy, O.G., Attia, A.M., Ghaithan, A.M., Mohammed, A. (2022). Spare parts supply with incoming quality control and inspection errors in condition based maintenance. Computers & Industrial Engineering, 172: 108534. https://doi.org/10.1016/j.cie.2022.108534

[3] Mzili, T., Mzili, I., Riffi, M. E., Pamucar, D., Simic, V., Abualigah, L., Almohsen, B. (2024). Hybrid genetic and penguin search optimization algorithm (GA-PSEOA) for efficient flow shop scheduling solutions. Facta Universitatis, Series: Mechanical Engineering, 22(1): 77-100. https://doi.org/10.22190/FUME230615028M

[4] Amery, R.W. (2023). Federal meat and poultry inspection duties and requirements-part 2: The public health inspection system, marks of inspection, and slaughter inspections. Journal of Environmental Health, 85(10): 16-19.

[5] Zhang, Z., Jain, A., Kumar, V. (2022). Model-based part manufacturing quality inspection path planning. Wireless Communications and Mobile Computing, 2022(1): 3119284. https://doi.org/10.1155/2022/3119284

[6] Volkau, I., Mujeeb, A., Dai, W., Erdt, M., Sourin, A. (2021). The impact of a number of samples on unsupervised feature extraction, based on deep learning for detection defects in printed circuit boards. Future Internet, 14(1): 8. https://doi.org/10.3390/fi14010008

[7] Djenouri, Y., Srivastava, G., Lin, J.C.W. (2024). Applied AI in defect detection for additive manufacturing: Current literature, metrics, datasets, and open challenges. IEEE Instrumentation & Measurement Magazine, 27(4): 46-53. https://doi.org/10.1109/MIM.2024.10540405

[8] Zhang, L., Dai, Y., Fan, F., He, C. (2022). Anomaly detection of GAN industrial image based on attention feature fusion. Sensors, 23(1): 355. https://doi.org/10.3390/s23010355

[9] Benzerrouk, S., Ludwig, R. (2007). Infrared detection of defects in powder-metallic compacts. Journal of Nondestructive Evaluation, 26(1): 1-9. https://doi.org/10.1007/s10921-007-0017-x

[10] Ullah, W., Khan, S.U., Kim, M.J., Hussain, A., et al. (2024). Industrial defective chips detection using deep convolutional neural network with inverse feature matching mechanism. Journal of Computational Design and Engineering, 11(3): 326-336. https://doi.org/10.1093/jcde/qwae019

[11] Movafeghi, A., Mohammadzadeh, N., Yahaghi, E., Nekouei, J., Rostami, P., Moradi, G. (2018). Defect detection of industrial radiography images of ammonia pipes by a sparse coding model. Journal of Nondestructive Evaluation, 37(1): 3. https://doi.org/10.1007/s10921-017-0458-9

[12] Niccolai, A., Caputo, D., Chieco, L., Grimaccia, F., Mussetta, M. (2021). Machine learning-based detection technique for NDT in industrial manufacturing. Mathematics, 9(11): 1251. https://doi.org/10.3390/math9111251

[13] Zeng, L., Liu, R., Xiong, L., Ho, J.C. (2025). TransFed: cross-domain feature alignment for semi-supervised federated transfer learning. Machine Learning, 114(8): 181. https://doi.org/10.1007/s10994-025-06805-1

[14] Chen, C., Wang, Y., Qin, Y., Tang, B. (2024). Low-priori intervention transfer learning network for large-span fault diagnosis. IEEE Sensors Journal, 24(20): 33455-33466. https://doi.org/10.1109/JSEN.2024.3448467

[15] Wang, H., Cheng, Y., Chen, C.P., Wang, X. (2021). Hyperspectral image classification based on domain adversarial broad adaptation network. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-13. https://doi.org/10.1109/TGRS.2021.3128162

[16] Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., Zhang, G. (2015). Transfer learning using computational intelligence: A survey. Knowledge-Based Systems, 80:

14-23. https://doi.org/10.1016/j.knosys.2015.01.010
[17] Klenk, M., Aha, D.W., Molineaux, M. (2011). The case for case-based transfer learning. AI Magazine, 32(1): 54-69. https://doi.org/10.1609/aimag.v32i1.2331
[18] Dornaika, F., Bi, J., Charafeddine, J., Xiao, H. (2025). Semi-supervised learning for multi-view and non-graph data using Graph Convolutional Networks. Neural Networks, 185: 107218. https://doi.org/10.1016/j.neunet.2025.107218
[19] Ziraki, N., Bosaghzadeh, A., Dornaika, F., Ibrahim, Z., Barrena, N. (2023). Inductive multi-view semi-supervised learning with a consensus graph. Cognitive Computation, 15(3): 904-913. https://doi.org/10.1007/s12559-023-10123-w
[20] Ramírez, F., Allende, H. (2013). Detection of flaws in aluminium castings: A comparative study between generative and discriminant approaches. Insight-Non-Destructive Testing and Condition Monitoring, 55(7): 366-371. https://doi.org/10.1784/insi.2012.55.7.366
[21] McKelvey, M.H., Babin, R.S. (1992). Engineering reliability and maintainability review-A regimen for discovering production deficiencies. In Annual Reliability and Maintainability Symposium 1992 Proceedings, Las Vegas, NV, USA, pp. 475-477. https://doi.org/10.1109/ARMS.1992.187867