This article has been accepted for publication in a future proceedings of this conference, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.1109/CVIP67348.2025.11291254,
2025 International Conference on Computer Vision, Image Processing and Computational Photography (CVIP)

High-frequency-aware Multi-task Learning and Transformation Consistency for semi-supervised Electromagnetic Shielding Optical Window Image Segmentation

1st Qiong Zhou
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
231612126@sust.edu.cn

2nd Tao Lei *
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
*leitao@sust.edu.cn

3rd Bo Liu
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
1114576740@qq.com

4th Tongfei Liu
Shaanxi Joint Laboratory of Artificial Intelligence
Shaanxi University of Science and Technology
Xi'an, China
liutongfei_home@hotmail.com

5th Yan Lu
College of Bioresources Chemical and Materials Engineering
Shaanxi University of Science and Technology
Xi'an, China
luyan@sust.edu.cn

6th Hongying Meng
The Department of Electronic and Electrical Engineering
Brunel University of London
London,U.K.
hongying.meng@brunel.ac.uk

***Abstract*—The electromagnetic shielding optical windows (OWs) are critical components in modern aircraft and electronic instruments. Accurate segmentation of crack template images can quantitatively analyze their structural parameters, thereby facilitating OWs design and manufacturing. However, pixel-level annotation is costly and labor-intensive. To address these problems, we propose a high-frequency-aware multi-task learning and transformation consistency for semi-supervised electromagnetic shielding optical window image segmentation network (HAMTC-Net) that enhances unlabeled data utilization through transformation consistency and high-frequency feature guidance. Specifically, CutMix and Mixup augmentations are incorporated to improve consistency regularization. An equivariant loss is introduced through an auxiliary classification task, which increases the global perception of the encoder. Furthermore, wavelet-based high-frequency features guide pixel-level consistency learning, enabling progressive learning from simple to complex patterns. Experiments on OWs crack image datasets demonstrate that HAMTC-Net outperforms existing state-of-the-art semi-supervised learning methods in segmentation accuracy.**

Keywords*—Optical Window Crack Image Segmentation, Semi-supervised Learning, Transformation Consistency, Multi-tasks, High-frequency-aware

I. INTRODUCTION

Metal mesh electromagnetic shielding optical windows (OWs) are designed to effectively block electromagnetic radiation, ensuring the stable operation of optoelectronic instruments and systems. They are widely used in precision electronics, aerospace, computing, and military applications [13].

As illustrated in Figure 1, the fabrication of OWs relies on random crack templates composed of grid regions and intervening cracks. Structural parameters such as the number, size, and distribution of grids, as well as crack width and duty cycle, which directly influence the final metal mesh structure after metallization, thereby affecting the electromagnetic shielding performance of the OWs [11]. Therefore, analyzing these parameters is essential for correlating fabrication conditions with OWs performance. To enhance the quality of OWs, researchers commonly utilize optical microscopy (OM) to image crack templates with random structures. The performance of OWs can be indirectly assessed by analyzing the structural parameters in these images. However, this analysis still primarily relies on manual measurement, which is time-consuming, inefficient, subjective, and lacks comprehensive statistical coverage. To address these limitations, we propose segmenting OM images of crack templates into binary maps, followed by digital image processing to enable rapid, objective, and accurate extraction of structural parameters.

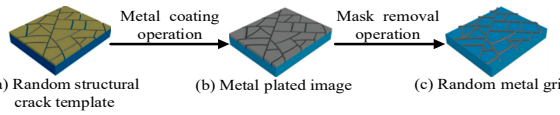

Figure 1. Schematic diagram of the preparation of metal mesh electromagnetic shielding.

Traditional image segmentation methods are time-consuming, labor-intensive with poor accuracy. The emergence of deep learning techniques has played a pivotal role in addressing such limitations. In particular, Convolutional Neural Networks (CNNs) have substantially advanced segmentation tasks. For example, U-Net proposed by Ronneberger et al. [12], greatly improves segmentation performance through its symmetric encoder-decoder architecture and has become a foundational model in the field. CNN-based methods effectively capture local features via hierarchical representations and are often enhanced by integrating modules such as pyramid feature fusion, attention mechanisms, and depth-wise separable convolutions, as seen in models like MFANet [6], SGU-Net [7], and DefED-Net [9].

Besides methods based on CNN, Transformer architectures have attained extraordinary success in natural language processing (NLP), and in recent times, they have also demonstrated considerable potential in image segmentation tasks [16]. The self-attention mechanism allows Transformer-based models to acquire global dependencies, a capability that enhances the accuracy and robustness of segmentation. Prominent models within this research field include Hresformer [2], and LW-CTrans [15].

Most modern segmentation networks depend heavily on large-scale labeled data to generalize effectively. However, annotating OM images of random crack templates is particularly labor-intensive, requiring expert-level, pixel-wise labeling, and is therefore time-consuming and costly. Meanwhile, ongoing OWs fabrication experiments are continuously producing large volumes of unlabeled OM data, posing a pressing need to exploit these data efficiently.

Recently, semi-supervised learning (SSL) has emerged as a promising paradigm in image segmentation, enabling competitive performance with limited labeled data [3,25]. Despite their success, existing SSL methods still face critical limitations. Firstly, many fail to fully exploit transformation consistency as an effective regularization strategy. Secondly, they often lack multi-task learning mechanisms to enhance the encoder's global representation capacity. Thirdly, they tend to overlook high-frequency information, which is crucial for capturing fine structural details and achieving robust generalization on large-scale unlabeled data.

To address the aforementioned challenges, we propose a high-frequency-aware multi-task learning and transformation consistency for semi-supervised electromagnetic shielding optical window image segmentation network (HAMTC-Net). Our major contributions are as follows:

(1) We construct a transformation consistency mechanism based on diverse image transformations to improve model robustness under complex scenes variations.

(2) We propose an equivariant loss and design a multi-task classifier after the encoder enabling the model to recognize image enhancement methods and improve the global learning ability of the model.

(3) We introduce a wavelet transform to extract high-frequency information and use it to screen pixel-level constrained unsupervised losses to focus on high frequency regions.

II. Related Work

A. Multi-tasks Based on Semi-supervised Learning

Multi-task learning integrates a variety of auxiliary tasks together with the core task of pixel-level segmentation, with the goal of improving the model's ability to represent shape and spatial features. Currently, there are two main types of frameworks in this field: the first is a parameter-sharing structure, which comprises a unified backbone network combined with multi-task output branches; the second involves building multiple independent networks, where each network is responsible for a different auxiliary task and operates with isolated parameters. For instance, Liu et al. [8] proposed a Glioma Multimodal MRI Analysis System (GMMAS) that employs a deep learning network to process multiple events at the same time, and it capitalizes on the interdependencies among these events through an uncertainty-based multi-task learning architecture. To tackle the problem of excessive interdependence between tasks, Liao et al. [5] introduced an Efficient Segmentation-Denoising Model (ESDM), a multi-task deep learning framework tailored for enhancing Optical coherence tomography (OCT) imaging. In addition, Lei et al. [25] presented a 3-D semi-supervised medical image segmentation framework where multi-consistency learning is embedded, serving the purpose of fuzzy perception-guided target selection. The core of this method is to enhance the high similarity between the prediction results produced by multiple perturbation networks for the same input image.

Although the aforementioned methods enhance local structural modeling, they still suffer from issues such as inter-task interference, limited global semantic representation, and inadequate robustness to perturbations introduced by image augmentation.

B. Consistency Learning in Semi-supervised Learning

Consistency learning is extensively employed in the SSL domain, primarily due to its capacity to fully utilize unlabeled data. Its core principle lies in maintaining consistent predictions for the same input after subjecting it to various perturbations. The Mean Teacher (MT) framework [14], which is a classic representative of this learning paradigm and is structured such that it applies data augmentation techniques (e.g., noise injection, rotation, scaling, etc.) to the student model and teacher model respectively. Moreover, consistency loss is utilized to drive the decision boundary toward low-density data regions, which in turn enhances the generalization capability of the segmentation model. Beyond this, Du et al. [24] present a semi-supervised medical image segmentation approach that is built on Collaborative intra-inter Contrastive Learning and Multi-Perspective Consistency (CCL-MPC). Li et al. [17] proposed a transformation-consistent self-ensembling model (TCSMv2), aiming to foster consistent predictions from the network being trained for the same input under various perturbations. Lei et al. [3] novel unified feature consistency (UFC) for semi-supervised medical image segmentation, which encompasses under-performing pixels (UPPs) and valid regions.The aforementioned methods improve unlabeled data

utilization by enforcing prediction consistency under perturbations but rely heavily on simple transformations and apply consistency mainly at the output layer, limiting adaptability to complex and diverse perturbations.

III. METHODS

A. Overview

As illustrated in Figure 2, the network architecture is based on the MT framework and comprises a supervised and an unsupervised branch. In the supervised branch, labeled data are fed into the student model. The cross-entropy (CE) loss and Dice loss are jointly used to form the supervised loss *Lsup*, which measures the discrepancy between the ground truth y^l and the prediction $\hat{y}^l$. The supervised loss is formulated as:

$$L_{sup} = 0.5 \times \left(CE(y^l, \hat{y}^l) + Dice(y^l, \hat{y}^l)\right) \quad (1)$$

For the unlabeled branch, the unlabeled data is first augmented via a transformation π_i and then passed through the student model. Simultaneously, the prediction from the teacher model undergoes the same transformation. This setup enforces consistency regularization, encouraging the model to produce consistent predictions under perturbations. Additionally, a high-frequency-aware strategy is employed during the early stage. It selectively filters high-frequency pixels to calculate the consistency loss L_c (see Section 3.2 for details). To further enhance the model's global representation capability, a transformation classifier is appended to the encoder of the student model. This auxiliary task introduces an equivariant loss L_e (see Section 3.3 for details). The overall unsupervised loss is defined as follows:

$$L_{unsup} = L_c + \gamma L_e \quad (2)$$

where γ = 0.1 in our work. The total objective function is defined as:

$$L = L_{sup} + \lambda(t) L_{unsup} \quad (3)$$

where λ(t) is a Gaussian warm-up function given by

$$\lambda(t) = e^{-5\left(1-\frac{t}{t_{max}}\right)^2} \quad (4)$$

where t denotes the current training step, and t_{max} represents the maximum number of training steps. As training progresses, the weight of the unsupervised loss gradually increases, reaching a maximum value of 1. The following subsections provide a detailed explanation of the three components of our proposed network.

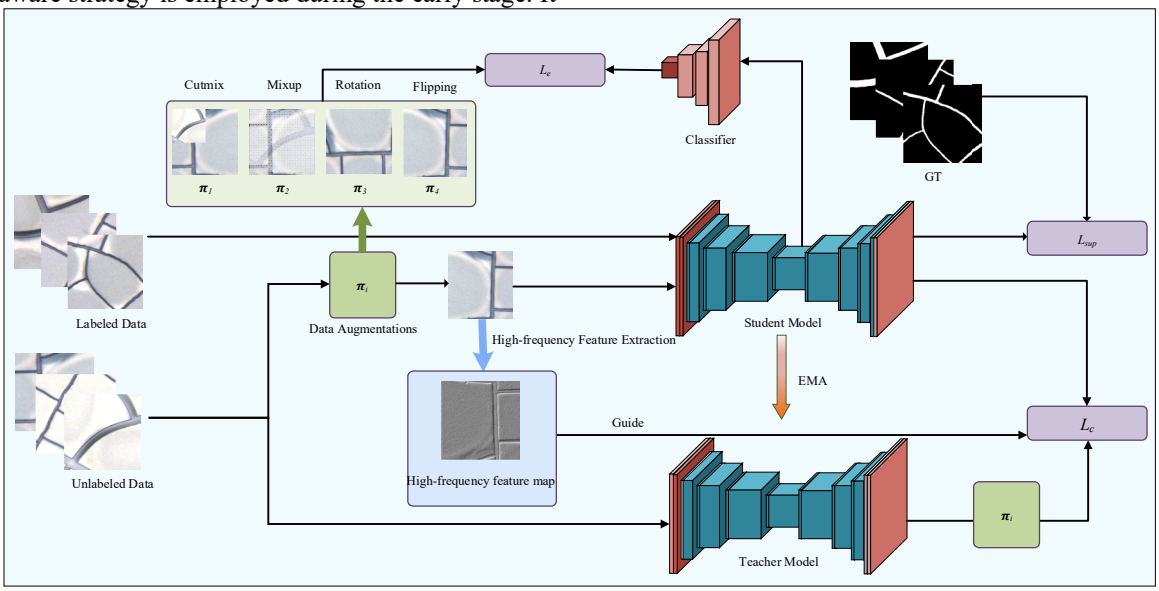

Figure 2. The overall architecture of HAMTC-Net network. The network consists of two branches: supervised and unsupervised. The supervised branch trains the student model on a small set of labeled data and supervised loss is L_{sup}. The unsupervised branch comprises three components: (1) applying four data augmentations π_i (CutMix, MixUp, rotation, and flipping) to unlabeled data, passing them through the student model, and using an auxiliary classifier after the encoder to predict the type of transformation which introduces an equivariant loss L_e; (2) performing wavelet decomposition to extract high-frequency features, which guide informative pixel selection and regularize the unsupervised loss, enhancing attention to high-frequency regions; (3) applying the same transformations to the teacher model's output to produce pseudo-labels, which combined with high-frequency guidance, are used to compute consistency loss L_c with the student predictions.

B. Full Augmentation-based Transformation Consistency

An ideal segmentation model should maintain output consistency under input perturbations. However, the geometric non-equivariance of convolutional features can cause spatial misalignment between teacher and student predictions after transformations, which can corrupt pseudo-labels and limit semi-supervised learning performance. To address this challenge, we propose a full augmentation-based transformation consistency method in the output space, which applies composite augmentations including CutMix [22], MixUp [23], rotation, and flipping differentially along the student and teacher paths and enforces output consistency, enabling the model to maintain stable predictions even under strong augmentations. For unlabeled data x_i, the student model first applies the transformation π_i to x_i, and the teacher model applies the same π_i after the output, and inserts random perturbation Gaussian noise in the process. The difference between the prediction results of the teacher model and the

student model is minimized by the mean square error loss (MSE). This approach not only improves the utilization of unlabeled data and ensures the reliability of pseudo-labels, but also guides the model to focus on global structure and shape information, significantly enhancing generalization and robustness. Through this mechanism, alignment bias caused by feature non-equivariance is effectively mitigated, and consistency regularization is transformed into a powerful representation learning tool. The loss term R is defined as follows:

$$R = ||z_i - \tilde{z}_i||^2 \tag{5}$$

where z_i and $\tilde{z}_i$ represent the predictions of the student and teacher models, respectively, defined as follows:

$$z_i = \theta^s\big(\pi_i(x_i)\big), \quad \tilde{z}_i = \pi_i\big(\theta^t(x_i)\big) \tag{6}$$

π_i denotes a set of random data augmentations applied to the input. θ^s and θ^t denote the student model and the teacher model, respectively.

To enhance the effect of consistency regularization in SSL, transformation consistency is extended to more complex types of augmentations.

C. Equivariant Loss

In strategies that build consistency through transformations, it is generally assumed that effective feature representations for segmentation should remain consistent under different transformations. This assumption holds for many simple augmentations. However, some transformations may conflict with the prior knowledge of segmentation tasks, for example, including geometric transformations and more complex strong augmentations. Therefore, the effective feature representations required for segmentation should exhibit equivariance or discriminability under different geometric transformations and strong augmentations. To address this issue, we introduce an equivariance loss into the segmentation model, as shown in Figure 2, the equivariant loss requires the encoder not only to predict the segmentation output but also to predict the type of augmentation applied to the input (CutMix [22], MixUp [23], rotation, and flipping). To correctly distinguish between different augmentation types, the model must understand the overall structure and semantic relationships of the input image, rather than relying solely on local textures or individual points. For example, rotation and flipping require the model to capture the orientation of the target, while CutMix [22] and MixUp [23] force the model to integrate global contextual information to identify the augmentation type. Therefore, the equivariant loss can indirectly promote the encoder to learn global semantic representations.

The segmentation model is defined in an encoder-decoder structure as follows:

$$f(x_i) = f_\gamma\big(f_\theta(x_i)\big) \tag{7}$$

where f denotes the entire segmentation network, f_θ is the encoder, and f_γ is the decoder.

For an input x_i, when it undergoes a transformation π_i, the corresponding segmentation result should also change accordingly.

$$f_\theta\big(\pi_i(x_i)\big) = \pi_i\big(f(x_i)\big) \tag{8}$$

Furthermore, it can be inferred that:

$$f_\theta\big(\pi_i(x_i)\big) \neq f_\theta(x_i) \tag{9}$$

Therefore, we enhance the learning of transformation information π_i within the encoder $f(\theta)$ by introducing a classifier $p_\emptyset(\cdot)$ to predict the type of random transformation applied. The equivariance loss L_e is defined as follws:

$$L_e = \frac{1}{C}\sum_{i=0}^{C-1} CE\left(p_\emptyset\left(f_\theta\big(\pi_i(x)\big)\right), i\right) \tag{10}$$

π_i represents one of four data augmentation types. Therefore, the number of categories is $C = 4$.

D. High-Frequency-aware Module

Curriculum learning is a training approach that simulates the human learning process by suggesting the model should begin with learning simple data and gradually move on to complex data [18]. [20] pointed out that most current methods ignore the frequency components of the image. And in the frequency domain, high-frequency components correspond to edges, textures, and fine details in the spatial domain. This regions are often error-prone in image segmentation tasks. Motivated by this principle, our approach begins by focusing on low-frequency content to stabilize early learning. As training progresses, high frequency components are gradually introduced to enhance the model's ability to delineate complex boundaries and fine structures.

Specifically, for each unlabeled input image, the student branch applies a transformation π_i, and the transformed image is then decomposed via wavelet transform to extract its high-frequency components. A consistency loss is subsequently computed under the guidance of these high-frequency maps, promoting attention to detailed regions throughout training. The specific process as seen follows:

As illustrated in Figure 3, the enhanced image is decomposed into low-frequency components, horizontal high-frequency components, vertical high-frequency components, and diagonal high-frequency components (denoted as LL, HL, LH, and HH) using wavelet transform [1]. L and H are defined as follows:

$$L = LL \tag{11}$$

$$H = HL + LH + HH \tag{12}$$

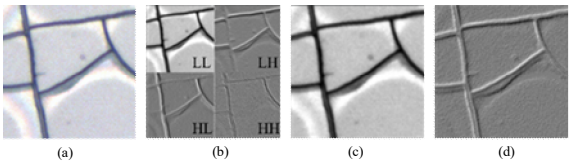

Figure 3. The results of wavelet decomposition of the OM image of the random crack template. (a) is the enhanced image; (b) shows the image after wavelet decomposition; (c) represents the low-frequency component L; (d) represents the high-frequency component H. It can be seen that H mainly focuses on the image details, whereas L emphasizes the image's overall semantic content.

Therefore, the high-frequency feature map H obtained from the wavelet decomposition is normalized pixel-wise as follows:

$$H_{norm}(x,y) = \frac{H_{(x,y)} - H_{min}}{H_{max} - H_{min}} \quad (13)$$

where x, y denote the pixel coordinates in the image, H_{min} and H_{max} represent the minimum and maximum pixel values in H, respectively, and H_{norm} is the normalized result.

After obtaining the normalized high-frequency map, unreliable predictions are filtered out and only reliable pixels are retained for student model training. The high-frequency-aware consistency loss L_c is designed as the pixel-wise mean squared error (MSE) loss between the teacher and student models:

$$L_c(\theta^t, \theta^s) = \frac{\sum_p I(u_p > LT) I(u_p < HT) R_p}{\sum_p I(u_p > LT) I(u_p < HT)} \quad (14)$$

$I(\cdot)$ denotes the indicator function. R_p represents the consistency result of the fully augmented prediction between the teacher and student models at pixel p. u_p denotes the value of the normalized high-frequency feature map at pixel p. HT and LT are the upper and lower threshold values, respectively.

After incorporating high-frequency-aware into training, the MT model can learn more reliable features and reduce overall uncertainty. LT and HT are dynamically adjusted by the Gaussian warm-up function as follows:

$$LT = LT_{max} - LT_{max} \cdot \lambda(t) \quad (15)$$

$$HT = HT_{max} + (1 - HT_{min}) \cdot \lambda(t) \quad (16)$$

λ(t) represents a Gaussian warming-up function. Regarding threshold setting, we referred to the empirical values used in existing semi-supervised segmentation methods and conducted multiple rounds of experimental tuning based on the characteristics of our task. We ultimately set $LT_{max} = 0.15$ and $HT_{min} = 0.65$, respectively. This configuration effectively balances the removal of low-confidence pseudo-labels with the retention of high-confidence predictions, thereby improving the quality of pseudo-labels and enhancing the stability of the training process. By applying the Gaussian warming-up function, the upper threshold is gradually increased to 1, while the lower threshold is progressively decreased to 0. In other words, as training progresses, our method gradually filters out less data, allowing the student model to learn from increasingly uncertain cases, starting from relatively confident ones.

IV. Experiments

A. Experimental Setup

The experiments were conducted using the PyCharm 2023.1 development environment and the PyTorch 1.11.0 deep learning framework. All training was performed on a GeForce RTX 3090 Ti GPU with 24GB of VRAM. During training, the Daubechies 2 (Db2) wavelet was used as the basis function for wavelet decomposition. The student model was optimized using the Adam optimizer with a learning rate of 0.001. The batch size was set to 8, and the total number of training epochs was 150.

B. Datasets

In our experiments, 88 OM images of randomly structured crack templates were collected using an Olympus BX53M optical microscope (resolution: 2592×1944) in the electromagnetic shielding window preparation laboratory. These images are characterized by a large number of grids of varying sizes, fine and curved cracks, and low contrast (see Figure 4 for examples). All original images were manually annotated in the LabelMe software to generate corresponding json files, which were then processed to produce labeled images with matching filenames. The original images are ultra-high resolution (2592×1944). During training, to fit the model input size, enhance data diversity, and prevent overfitting, each image is cropped into 256×256 patches. On average, each original image yields about 76 patches, totaling approximately 6,688 samples. The dataset was segmented into 27 training images and 66 test images, and every image underwent cropping to a uniform size of 256×256 pixels. The training set was further split into labeled and unlabeled subsets, with the division ratios set to [5%, 95%] and [10%, 90%] for the two subset types respectively.

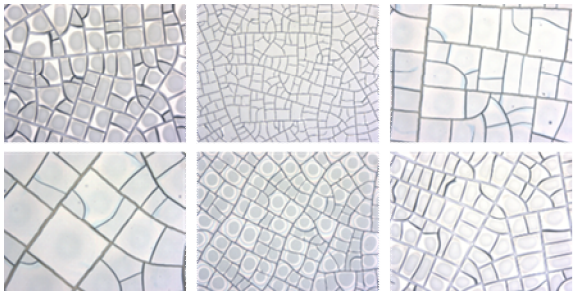

Figure 4. The Random structural crack template OM image dataset.

C. Evaluation Metrics

The segmentation results of different algorithms are evaluated using commonly adopted metrics in mainstream image segmentation literature, including the Dice coefficient (DI), Jaccard index (JA), and Sensitivity (SE).

(1) The Dice coefficient (DI) primarily measures the similarity between two seg mentation maps and is one of the main evaluation metrics in the field of medical image segmentation. Its formula is as follows:

$$DI = \frac{2 \times TP}{FP + 2 \times TP + FN} \quad (17)$$

(2) The Jaccard similarity coefficient (JA), also known as the Intersection over Union (IoU), represents the degree of overlap between samples. The formula is as follows:

$$JA = \frac{TP}{FP + TP + FN} \quad (18)$$

(3) Sensitivity(SE) represents the ability of the algorithm to correctly predict positive cases. The formula is expressed as follows:

$$SE = \frac{TP}{TP + FN} \quad (19)$$

D. Experimental Comparision and Analysis

1) Comparative Experiments

To validate the effectiveness of the proposed method, we compare it with several mainstream SSL approaches, including MT [14], UA-MT [21], TCSMv2 [17], CPS [4], DTC [10], and MC-Net [19]. The experimental results are presented in Table 1, where our method consistently achieves superior performance across different proportions labeled data. Specifically, when

using only 5% of the ground truth labels, our method surpasses TCSMv2 [17] method by 0.88% in terms of Dice coefficient. With 10% labeled data, our method still outperforms TCSMv2 [17] by 0.7% in the Dice coefficient, demonstarting its robustness and effectiveness under low-label regimes.

2) Ablation Studies

To further verify the effectiveness of each module in the proposed method, we conduct ablation studies on the OM crack image dataset, using only 5% of the labeled data randomly sampled from the full dataset. The experimental results are shown in Table 2.

Without any semi-supervised techniques, the baseline segmentation network U-Net, trained with only 5% of the labeled data (denoted as SupOnly), achieves a Dice score of 81.20%. To leverage unlabeled data and enhance the encoder's global feature learning capability, we evaluate the following strategies.

TABLE I. TABLE TYPE STYLES STATISTICAL COMPARISON WITH SOTA METHODS ON THE RANDOM CRACK TEMPLATE OM DATASET. (NOTATIONS:THE BEST VALUES ARE IN BOLD.)

Methods	**5% Labeled Data**				**10% Labeled Data**			
	Label	***Dice(%)***	***JA(%)***	***SE(%)***	***Label***	***Dice(%)***	***JA(%)***	***SE(%)***
SupOnly	1/27 5%	81.20	70.58	79.04	2/27 10%	83.91	74.04	83.47
MT [14]		84.22	74.38	80.26		85.41	76.69	86.92
UA-MT [21]		85.76	76.81	86.71		87.70	79.79	87.30
TCSMv2 [17]		86.66	78.32	**89.99**		88.86	80.98	88.30
CPS [4]		85.38	76.30	87.85		87.24	78.43	86.11
DTC [10]		84.62	75.36	84.12		85.26	77.56	86.47
MC-Net [19]		85.37	77.61	86.47		87.38	78.74	83.92
HAMTC-Net		**87.54**	**78.78**	87.30		**89.56**	**82.07**	**89.76**
FullSupervised	100%	92.09	86.13	93.15	100%	92.09	86.13	93.15

Firstly, building on the MT framework, transformation π_i and equivariant loss L_e are introduced (Scheme 2). Secondly, to enable the student model to learn from easy to complex data and gradually learn from more reliable targets, a high-frequency-aware module is introduced (Scheme 3). Then, based on Scheme 3, the equivariant loss L_e is further integrated (Scheme 5). Finally, to improve the efficiency in utilizing unlabeled data, transformation consistency is applied either by applying a shared transformation π_i to the teacher model's output (Scheme 1), or by combining transformation consistency with the two proposed modules to validate the overall effectiveness of the proposed approach (Schemes 4, 6, and Ours).

a) Ablation study on Transformation Consistency.

By comparing the results of Schemes 1, 2, and 3, it can be observed that transformation consistency, when used independently, provides the most significant improvement in SSL performance compared to the other two modules. Specifically, it improves the Dice coefficient by 2.23% over the MT baseline. This indicates that introducing more complex transformations, such as CutMix [22] and Mixup [23] (denoted as π_i), helps the network extract more robust and discriminative features from the samples, thereby improving segmentation performance.

b) Ablation study on Equivariant Loss and High-frequency-aware Module.

As shown in Table 2, Scheme 5, which integrates equivariant loss L_e and high-frequency-aware module, improves the Dice score from 86.23% (Scheme 2) to 86.92%. This indicates that high-frequency-aware module complements L_e by mitigating the potential over-strength of random transformations π_i, which might otherwise hinder the student model's learning. Furthermore, comparing Scheme 1 and Scheme 6, introducing an auxiliary classification task to implement Le boosts the Dice score from 86.45% to 87.16%. This demonstrates that Le raises the encoder to recognize which transformation π_i has been applied, enhancing the model's global perception. While transformation consistency constrains the model's predictions, L_e directly optimizes the encoder parameters, forming a complementary mechanism that further boosts segmentation performance.

In the high-frequency-aware module, the choice of wavelet basis functions is crucial for effective feature decomposition. Given the wide variety of wavelets, we selec-

TABLE II. ABLATION STUDIES OF EACH COMPONENT. (NOTATIONS:THE BEST VALUES ARE IN BOLD.)

Methods	**Transformation Consistency**	**Equivariant Loss**	**High-frequency-aware Module**	**Dice(%)**
SupOnly				81.20
Scheme1	✓			86.45
Scheme2		✓		86.23
Scheme3			✓	85.78
Scheme4	✓		✓	87.11
Scheme5		✓	✓	86.92
Scheme6	✓	✓		87.16
Ours	✓	✓	✓	**87.54**

ted three representative and widely used orthogonal functions for comparison: Haar, sym2, and Db2. Haar serves as the simplest baseline, sym2 provides improved symmetry and smoothness, and Db2 offers higher smoothness and better time-frequency localization. As shown in Table 3, Db2 outperforms Haar and sym2 by 2.56% and 1.88%, respectively, confirming its superiority in modeling high-frequency features. Therefore, Db2 is adopted in our method.

TABLE III. ABLATION STUDIES ON DIFFERENT WAVELET BASIS FUNCTIONS IN THE HIGH-FREQUENCY-AWARE MODULE. (NOTATIONS:THE BEST VALUES ARE IN BOLD.)

Wavelet Basis Functions	Dice(%)
Haar wavelet	84.98
Symlets 2 wavelet (sym2)	85.66
Daubechies 2 wavelet (Db2)	**87.54**

3) Visualization Results

To intuitively demonstrate the effectiveness of HAMTC Net, Figure 5 shows segmentation results on the crack template OM image test set with only 10% of the data labeled. As shown, HAMTC-Net produces more accurate predictions with clearer boundaries, closely matching the ground truth. This is attributed to the integration of advanced transformation consistency and an equivariant loss function, which jointly enhance feature representation. In particular, for small light spots in random template images, the high-frequency-aware module effectively captures high-frequency details and guides the model to focus on more informative regions, thereby improving segmentation accuracy.

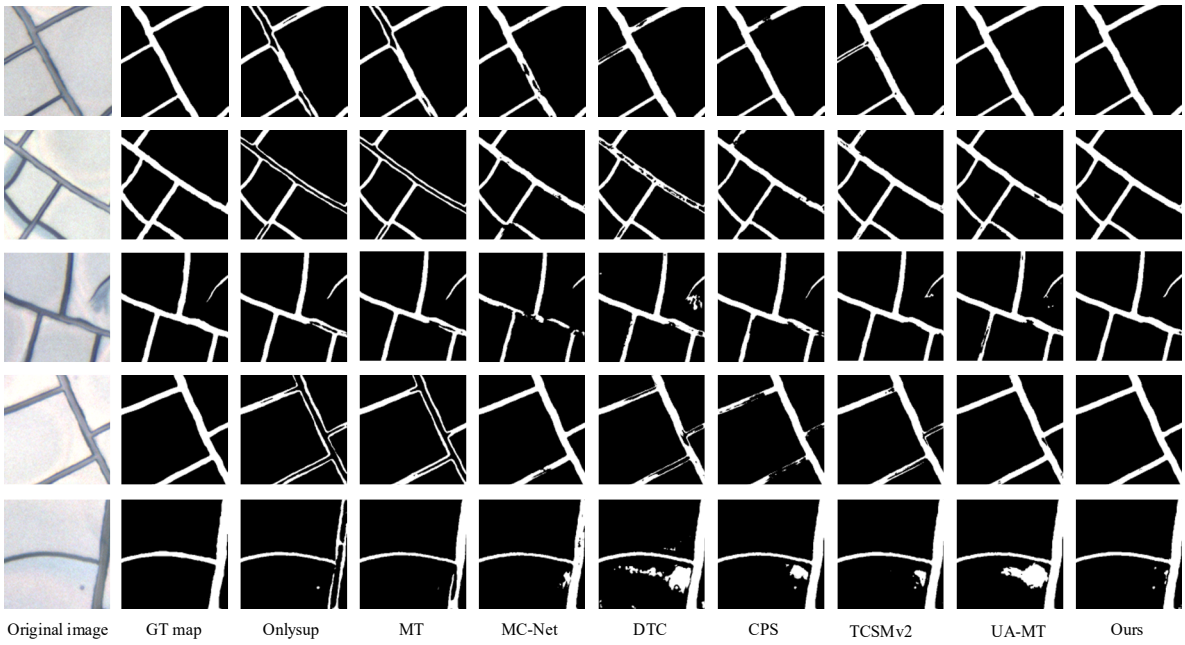

Figure 5. Visualization results compared with SOTA semi-supervised learning and supervised methods with 10% labeled data.

V. CONCLUSION

In this paper, we propose a new SSL approach for image segmentation of OWs. The proposed framework incorporates advanced transformation consistency constraints, an equivariant loss function along with an auxiliary classification task, and a wavelet-driven high-frequency- aware module. This integrated strategy boosts both global feature learning and pixel-level supervision, greatly enhancing segmentation performance on the OM image dataset and providing a solid basis for research into electromagnetic shielding OWs.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (Program No. 62271296 , 62201334), in part by The Innovation Capability Support Plan Project in Shaanxi Province (2025RS-CXTD-012), and in part by Scientific Research Program Funded by Shaanxi Provincial Education Department (23JP014, 23JP022).

REFERENCES

[1] E. M. Shalby, A. Y. Abdelaziz, E. S. Ahmed, *et al.*, "A comprehensive guide to selecting suitable wavelet decomposition level and functions in discrete wavelet transform for fault detection in distribution networks," *Scientific Reports*, vol. 15, no. 1, p. 1160, 2025.

[2] S. Ren and X. Li, "Hresformer: Hybrid residual transformer for volumetric medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[3] T. Lei, Y. Wang, X. Wang, Y. Wang, X. Wang, B. Hu, and A. K. Nandi, "Unified feature consistency of under-performing pixels and valid regions for semi-supervised medical image segmentation," *IEEE Trans. Radiat. Plasma Med. Sci.*, 2024, doi: 10.1109/TRPMS.2024.3465561.

[4] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2613–2622, 2021.

[5] J. Liao, T. Zhang, S. Shepherd, *et al.*, "Semi-supervised assisted multi-task learning for oral optical coherence tomography image segmentation and denoising," *Biomedical Optics Express*, vol. 16, no. 3, pp. 1197–1215, 2025.

[6] Y. Zhang, J. Yin, Y. Gu, *et al.*, "Multi-level Feature Attention Network for medical image segmentation," *Expert Systems with Applications*, vol. 263, p. 125785, 2025.

[7] T. Lei, R. Sun, X. Du, H. Fu, C. Zhang, and A. K. Nandi, "SGU-Net: Shape-guided ultralight network for abdominal image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1431–1442, 2023.

[8] Y. Liu, Z. Cui, L. Li, *et al.*, "Glioma multimodal MRI analysis system for tumor layered diagnosis via multi-task semi-supervised learning," *arXiv preprint*, arXiv:2501.17758, 2025.

[9] T. Lei, R. Wang, Y. Zhang, Y. Wan, C. Liu, and A. K. Nandi, "DefED-Net: Deformable encoder-decoder network for liver and liver-tumor segmentation," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 1, pp. 68–78, Jan. 2022.

[10] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 8801–8809, 2021.

[11] L. Yang, Y. Guan, F. Gao, *et al.*, "Effect of crack template structure morphology on electromagnetic shielding efficiency and visual performance of metal mesh optical window," *Results in Engineering*, vol. 25, p. 103888, 2025.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*, vol. 18, pp. 234–241, Springer, 2015.

[13] D. Tan, C. Jiang, Q. Li, S. Bi, X. Wang, and J. Song, "Development and current situation of flexible and transparent EM shielding materials," *Journal of Materials Science: Materials in Electronics*, vol. 32, pp. 25603–25630, 2021.

[14] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[15] H. Kuang, Y. Wang, X. Tan, *et al.*, "LW-CTrans: A lightweight hybrid network of CNN and Transformer for 3D medical image segmentation," *Medical Image Analysis*, vol. 102, p. 103545, 2025.

[16] S. S. Kumar, "Advancements in medical image segmentation: A review of transformer models," *Computers and Electrical Engineering*, vol. 123, p. 110099, 2025.

[17] X. Li, L. Yu, H. Chen, *et al.*, "Transformation-consistent self-ensembling model for semi-supervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 523–534, 2020.

[18] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4555–4576, 2021.

[19] Y. Wu, M. Xu, Z. Ge, J. Cai, and L. Zhang, "Semi-supervised left atrium segmentation with mutual consistency training," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II*, vol. 24, pp. 297–306, Springer, 2021.

[20] D. Xue, T. Lei, S. Yang, Z. Lv, T. Liu, Y. Jin, and A. K. Nandi, "Triple change detection network via joint multifrequency and full-scale Swin-transformer for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[21] L. Yu, S. Wang, X. Li, C. W. Fu, and P. A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II*, vol. 22, pp. 605–613, Springer, 2019.

[22] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6023–6032, 2019.

[23] H. Zhang, M. Cisse, Y. N. Dauphin, *et al.*, "mixup: Beyond empirical risk minimization," *arXiv preprint*, arXiv:1710.09412, 2017.

[24] X. Du, Y. Zou, T. Lei, *et al.*, "CCL-MPC: Semi-supervised medical image segmentation via collaborative intra-inter contrastive learning and multi-perspective consistency," *Neurocomputing*, vol. 621, p. 129287, 2025.

[25] T. Lei, W. Song, W. Zhang, X. Du, C. Li, L. He, and A. K. Nandi, "Semi-supervised 3D medical image segmentation using multi-consistency learning with fuzzy perception-guided target selection," *IEEE Trans. Radiat. Plasma Med. Sci.*, 2024, doi: 10.1109/TRPMS.2024.3473929.