Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **4**, December **2025**

Marine Raft Extraction Using Non-Salient Feature and Boundary Weight in Remote Sensing Images

Jiahua Wan[1,2,*], Vladimir Y. Mariano[1]

[1]College of Computing and Information Technologies,
National University, Manila 1008, Philippines
[2]Big Data and Artificial Intelligence College,
Anhui Xinhua University, Hefei 230088, P. R. China
wanjiahua@axhu.edu.cn, vymariano@national-u.edu.ph

*Corresponding author: Jiahua Wan

Received June 19, 2025, revised August 03, 2025, accepted August 05, 2025.

Abstract. *While existing convolutional neural networks (CNNs) can rapidly and accurately extract marine targets from remote sensing imagery, identifying marine rafts remains challenging due to complex coastal terrain, extensive shorelines, diverse raft morphologies, dynamic movements, and cluttered backgrounds. Furthermore, repeated pooling in CNNs causes spatial information loss, leading to inaccurate raft boundary segmentation, while limited convolutional receptive fields result in discontinuous segmentation. To address these limitations, this paper proposes three novel modules: a Local Expansion Attention (LEA) module to establish dependencies between adjacent feature regions and alleviate discontinuity in large raft segmentation; a Non-Salient Feature Extraction (NSF) module to enhance identification of rafts with background-similar features; and a Binary Cross-Entropy Loss with Boundary Weight Enhancement (BCE-BWE) module to improve boundary focus without increasing complexity. These modules are integrated into WNUNet, an efficient UNet-based model. Comparative experiments demonstrate the effectiveness of LEA, NSF, and BCE-BWE. On the constructed WJD remote sensing dataset, WNUNet achieves state-of-the-art performance with 90.91% IoU and 95.24% F1-score, outperforming existing raft extraction methods.*
Keywords: Non-salient feature; Boundary weight; Attention module; Remote sensing image; Marine Raft Extraction.

1. **Introduction.** The growing demand for marine products and economic incentives in China have spurred disorderly and illegal aquaculture practices, resulting in significant environmental degradation and socioeconomic losses. This poses substantial challenges to integrated coastal zone management [1, 2]. Consequently, rapidly and accurately identifying marine aquaculture types, mapping their spatial distributions, and dynamically monitoring raft aquaculture areas are critical for effective coastal resource management and sustainable ecological development [3, 4].

Remote sensing technology significantly enhances large-scale observation capabilities in complex, multi-scale marine environments, providing a cost-efficient solution that surpasses traditional field mapping and chart comparison methods [5]. Over the past decade, object-based image analysis (OBIA) has emerged as the predominant approach, segmenting images prior to classifying objects using patch-level features (e.g., spectral, textural, and structural characteristics) to achieve robust classification accuracy [6]. Nevertheless, raft aquaculture extraction faces persistent challenges: high-resolution imagery reveals significant structural and textural variations in rafts due to farming density and zone-specific characteristics. Moreover, as rafts are partially submerged, their spectral signatures and orientations are highly susceptible to dynamic environmental conditions such as wave action and water turbidity.

Advances in remote sensing have expanded its applications while improving spatial resolution, enabling efficient large-scale monitoring of floating rafts through image interpretation. Accordingly, accurate raft

extraction has become an active research focus. However, two major obstacles persist: (1) complex background clutter containing non-raft objects (e.g., sea dams, cages, trawls) exhibits spectral and textural similarities to rafts [7], and (2) rafts themselves demonstrate substantial appearance variations, scale diversity, and irregular geometries [8]. These factors collectively impede precise extraction.

Current raft extraction methodologies fall into two categories: traditional image processing and deep learning-based approaches. Traditional methods rely on manually engineered algorithms leveraging geometric, textural, and spectral features. Such approaches are vulnerable to interference from raft shape heterogeneity and complex backgrounds, often requiring scenario-specific feature engineering. Consequently, they exhibit limited scalability for large-area, multi-scene applications while demanding substantial manual intervention. In contrast, deep learning-based methods employ convolutional neural networks to autonomously learn discriminative raft features from extensive training data. This paradigm enables streamlined algorithm development for large-scale, complex coastal imagery while maintaining strong generalization capabilities [9].

2. **Objectives.** Semantic segmentation of remote sensing images can achieve accurate segmentation of floating rafts. Semantic segmentation is a pixel-level classification task, that can be used to comprehensively interpret remote sensing images and is an important means of decoding remote sensing images [10]. After proposing the fully convolutional network (FCN) [11] and laying out the network structure of the codec, researchers have designed various end-to-end image semantic segmentation networks. Among them, the ASPP of DeeplabV3+ [13], the Feature Pyramid Network (FPN) [14], the skip connection of UNet [12], and the parallel structure of HRNet [15] are models that can effectively solve the multi-scale problem of floating rafts. However, some researchers believe that the receptive field of convolution can only capture local context information, but cannot build global context information, which is very important for semantic segmentation. At the same time, downsampling the image through the model leads to the loss of spatial information on the feature map, and the model cannot accurately identify the boundary of the object.

To further improve the segmentation of remote sensing images using semantic segmentation networks, researchers have begun to use the attention mechanism to construct the global context information of the image. Among these works is a self-attention model based on transformers [16, 17, 18, 19]. For example, SETR [20] converts images into sequences and learns the semantic features of the images in sequence-to-sequence mapping. This mapping method not only constructs global context information, but also avoids the loss of spatial information during image downsampling. In addition, considering that the sequence-to-sequence mapping method brings huge complexity to the model, researchers have begun to design lightweight attention modules according to the principle of self-attention [21]. The attention module and convolution are combined to construct a semantic segmentation network. This enables the semantic segmentation network to build long-term dependency information while learning local context information. For example, DANet [22] is a semantic segmentation model built using ResNet [23] with a channel attention module and a spatial attention module. ResNet is used to learn the deep features of the image, the attention module is used to build the dependency information between each position and the global position, and feature concatenation is performed to enable the network to extract local and global context information. However, the attention module in DANet still introduces $O(N^2)$ time and space complexity to the network ($N = H \times W$, H and W denote the height and width of the image, respectively), which increases the difficulty of processing large-scale remote sensing images. Therefore, researchers designed lightweight attention modules. For example, channel attention modules, such as SE [24], SESR [25], and ECA [26], reduce the computational complexity of learning channel weights by compressing spatial information. CBAM [27] reduces the computational complexity of learning spatial location weights by compressing the channel, whereas modules such as CA [28] and SA [29] use local compression to simultaneously learn the weights in the channel and the spatial dimension. These lightweight attention modules can be randomly embedded anywhere in a semantic segmentation network without imposing a large burden on the network. Therefore, the attention mechanism has been widely used in semantic segmentation research to further improve the segmentation performance of semantic segmentation networks.

Based on the efficiency and attention mechanism of semantic segmentation networks, they are widely used in remote sensing image extraction. For example, MAP-Net [30] constructs multi-branch parallel paths to preserve accurate spatial information and learn multi-scale features. Subsequently, the map network uses the channel attention module and pyramid pooling module to enhance the fusion of multi-scale features, which improves the ability of the network to extract floating rafts from remote sensing images. MSRF-Net [31] uses convolution at different scales to build multi-scale spatial attention, and

then builds multi-scale channel attention through max pooling at different scales, to build a multi-scale receptive field encoder and a multipath decoder, thereby improving the ability of the model to learn multi-scale semantic features. B-FGC-Net [32] uses a residual unit and spatial attention module to build an encoder to highlight the spatial information of the feature map. In the decoder part, the global feature information perception module is used to capture multi-scale context information and integrate global information. MSL-Net [33] uses an inverted residual structure and an ASPP module to build efficient multi-scale context information and uses deformable convolution to enhance the ability of the model to identify irregularly shaped floating rafts. AFL-Net [34] designed an attention multi-scale feature (AMFF) module and a shape feature refinement (SFR) module to improve the accuracy of marker recognition in complex environments. MHA-Net [35] used multi-path hybrid dilated convolution and dense upsampling convolution to aggregate rich multi-scale features, and then used the channel attention module to reduce the noise in the features to improve the network's ability to extract small objects. Moreover, given the limitations of convolutional receptive fields, researchers have begun to apply vit [36] to the field of correlation extraction. Examples include BuildFormer [37], DSFENet [38], and Swi-U-S [39]. All utilize the ability of Vit to construct global contextual information of feature maps, which enables the model to learn more salient object features, thus improving the accuracy of the model in extracting floating rafts. However, the semantic segmentation model is a pixel-by-pixel classification prediction, and when extracting floating rafts from remote sensing images, it can easily lead to discontinuous or empty extraction of floating rafts owing to the large ocean area. Moreover, although the above deep learning-based models easily learn the salient features of floating rafts, it is difficult to identify some floating rafts with similar properties in the background, which may lead to a situation in which some floating rafts cannot be recognized, or the background is considered a floating raft.

Accurate identification of the boundary information of floating rafts is another key factor in describing their integrity of floating rafts [40]. However, the semantic segmentation model can quickly and accurately extract most offshore facilities. However, this model could not fully identify the boundaries of some irregular floating rafts. This is because in the codec model, to reduce the computational cost and increase the receptive field of the convolution, the model downsamples the image to a smaller-scale feature map and then restores it to the size of the original image after upsampling. This process may lead to a loss of boundary information. Additionally, the boundaries of floating rafts are susceptible to shadows, excessive marine algae, and other complex background objects. Therefore, the complete extraction of the floating raft boundary information is more challenging. Many models incorporate skip connections between the encoder and decoder to enhance their ability to extract object boundaries [41, 42, 43]. The skip connection can introduce the spatial information of the shallow feature map in the encoder into the decoder so that the decoder has accurate spatial information to accurately identify the boundary structure of the target. However, the skip connection also transfers the fuzzy semantic information in the shallow feature map to the decoder layer, thereby increasing the semantic noise in the decoder layer and affecting the segmentation results of the model. To better extract boundary information, researchers have introduced boundary supervision modules to prompt feature maps to learn the boundary features. For example, BESNet [44] selects multiple feature maps at different scales from the encoder and feeds them into the boundary feature learning module, which performs upsampling, convolution, and concatenation to generate new feature maps. These new feature maps are then supervised by sea surface true boundary labels to guide the learning of boundary features. Finally, A feature map containing boundary features is integrated into the decoder to improve the ability of the model to identify the target boundary. HED [45] performs boundary prediction on each layer of the feature maps in the encoder, supervised by the sea surface true boundary label loss. The predicted results are then concatenated and fused to generate feature maps that are mapped into high-dimensional features. This directly guides the object segmentation results to achieve accurate boundary generation. Both [46] and [47] introduced learning modules that incorporated relevant boundary information into the segmentation model and ensured that the output of these boundary modules was supervised by true boundary labels of the sea surface. Although the boundary feature learning module supervised by the true boundary labels of the sea surface can improve the ability of the model to recognize boundary features, it also incurs significant computational costs as an additional branch module in the model.

In this study, a WNUNet network is proposed to solve the above problems. To address the problem of discontinuous raft extraction, WNUNet incorporates a local expansion attention (LEA) module to aggregate the relationships between groups of neighboring pixels. To solve the misclassification of floating rafts with similar characteristics to the background, WNUNet introduces a non-salient feature extraction (NSF) module, which suppresses salient features and uses non-salient features to identify floating rafts with similar characteristics to the background. To overcome the problem of fuzzy construction boundary

identification, this study proposed a binary cross-entropy loss function with boundary weight enhancement (BCE-BWE). This loss function improves the accuracy of the constructed boundary extraction and reduces the high computational cost caused by a similar boundary feature learning branch module in the model. Therefore, this study makes the following contributions.

- In this paper, we propose a local expansion attention (LEA) module. By locally compressing and mapping feature maps, the LEA module enables the model to learn high-dimensional features while establishing local spatial dependencies, thus helping to segment complete floating rafts.
- In this study, we designed a non-salient feature extraction (NSF) module. By properly suppressing the salient features of the floating raft, the NSF module highlights the non-salient features of the floating raft and uses these features to capture parts of the floating raft that are difficult for the NSF module to identify.
- A binary cross-entropy loss function with boundary weight enhancement (BCE-BWE) was introduced to enhance the attention of the model to building boundary information.
- Using the LEA module, NSF module, and UNet, an efficient floating raft extraction model, WNUNet was constructed. Under the supervision of BCE-BWE, WNUNet achieved significant segmentation results for the constructed WJD dataset.

3. Methodology.

3.1. **Architecture of WNUNet.** The UNet model is widely used in binary image segmentation tasks and exhibits good segmentation performance. Each layer of the encoder in the UNet model consisted of two 3×3 convolutions and a Max pooling module. The encoder part consists of skip connections, upsampling, and two 3×3 convolutions. The key advantage of the UNet model is its skipped connection. These connections bring spatial information from shallow feature maps to the decoder, which helps the decoder recover precise spatial details by exploiting the rich spatial information in the shallow feature maps. This operation not only solves the loss of spatial information in the max-pooling process but also alleviates the problem of multi-scale data. However, in UNet, shallow feature maps have lower semantic information and smaller receptive fields owing to the convolution operation. The direct fusion of shallow feature maps and high-level semantic deep feature maps introduces semantic noise to the latter, which affects segmentation results. Furthermore, as a deep convolutional model, UNet tends to learn salient features about the target object, but has difficulty identifying objects with features similar to the background or misclassifying them as background.

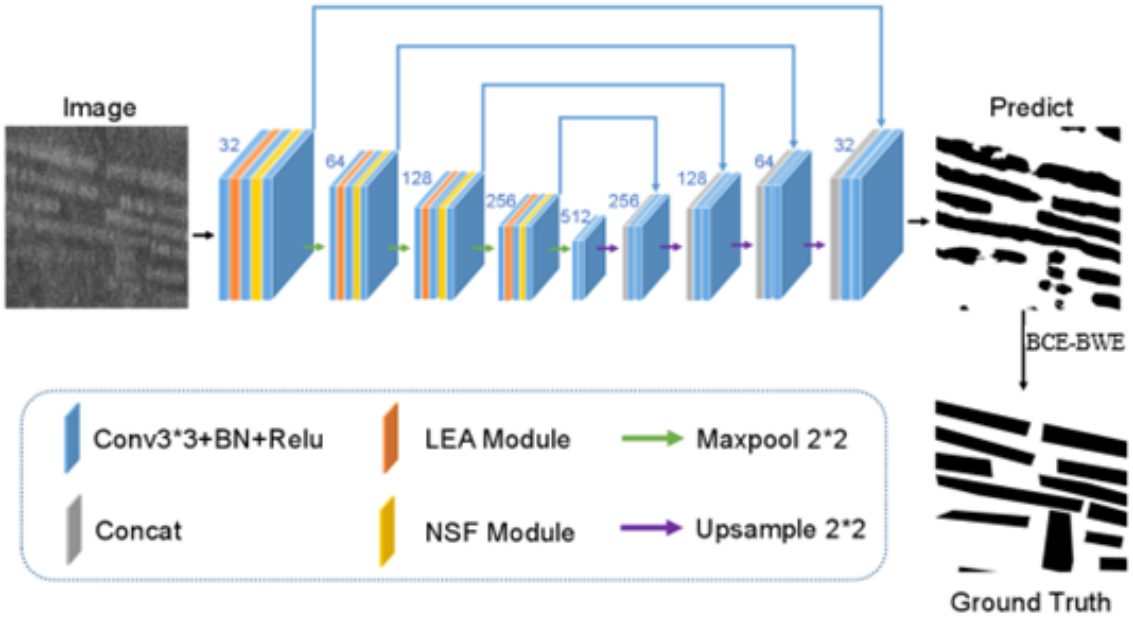

FIGURE 1. Structure of WNUNet.

To further understand the floating raft features in remote sensing images, the local expansion attention (LEA) module and non-salient feature extraction (NSF) module were integrated into the encoder of the UNet model to create the WNUNet model, as shown in FIGURE 1. In the encoder part of WNUNet,

the order of each layer structure is convolution: 3×3, LEA module, 3×3 convolution, NSF module, and another 3×3 convolution. A 3×3 convolution is used to extract features related to floating rafts in the image. The LEA module builds local spatial information and global channel dependencies in the feature map to enhance the continuity of the object segmentation. The NSF module extracts the non-salient features of floating rafts to help the model learn more valuable floating raft features. To prevent the LEA and NSF modules from interfering with each other, a 3×3 convolution was added between the two modules. Finally, this study used the Binary Cross-entropy loss function with boundary weight enhancement (BCE-BWE) to supervise the WNUNet model, aiming to improve the model's ability to identify floating raft boundaries. WNUNet utilizes the advantages of skip connections in UNet, and solves the above shortcomings of the UNet architecture by embedding LEA and NSF modules, thus enhancing the ability of the model to extract floating rafts from ocean remote sensing images.

3.2. **Local expansion attention (LEA) module.** To improve the continuity of the construction extraction in semantic segmentation models, we propose a local extended attention module. Attention modules are widely used in building extraction tasks to reduce semantic noise and capture local and global context information. However, most attention modules focus only on a single dimension of a channel or space, limiting their effectiveness in semantic segmentation models. To address this limitation, we constructed a locally extended attention module that focuses on both channel and spatial dimensions. This module aims to establish a relationship between a local region and its adjacent parts in an ocean remote sensing image, to promote the continuity of floating raft segmentation within the model.

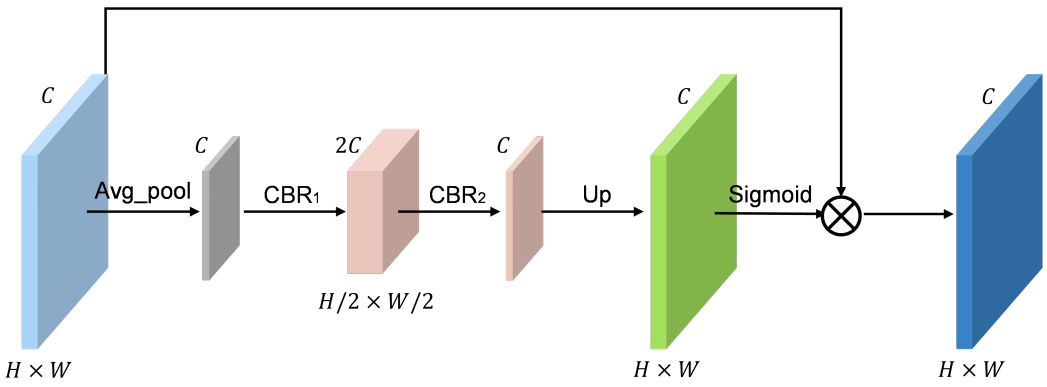

FIGURE 2. Local expansion attention (LEA) module. "Avg_pool" means Average Pooling. "CBR" means Conv + BatchNormalization + Relu. "Up" means upsampling. "⊗" means element-wise product.

The LEA module is illustrated in FIGURE 2. Inspired by the SE module, we constructed a LEA module. In the SE module, global average pooling is applied to the feature maps, leaving only the channel dimensions for attention. Then, two fully connected layers and a nonlinear activation function were used to calculate the weight relationship between the channels globally. Finally, the SE module learns the channel weights and applies them to the original feature maps via multiplication. Although the SE module establishes global dependencies along the channel dimension, it ignores the information in the spatial dimension. In contrast, the proposed LEA module replaces the global average pooling operation in the SE module with a 2×2 average pooling, which preserves the spatial information of the feature maps and lays the foundation for constructing local spatial dependencies. Specifically, we first perform 2×2 average pooling on the feature maps and then use two 3×3 convolutions with nonlinear activation functions to learn local spatial relationships and global channel relationships. Finally, through upsampling and sigmoid, the learned channel and spatial relationships were mapped to weights corresponding to the input. This process can be described as follows.

$$\omega = \sigma(Up(CBR_2(CBR_1(X)))) \tag{1}$$

$$output = multiply(input, \omega) \tag{2}$$

In equation (1), X represents the 2×2 average pooling feature map, whose shape size is compressed to $1/2$ of the original feature map. $CBR1$ and $CBR2$ represent the first and second layers of 3×3 convolution, followed by batch normalization and ReLU, respectively. σ is the sigmoid function. Multiply by indicates

the element-wise product. The difference between the first and second convolutional layers is that the first layer maps the feature map from the C channel to the $2C$ channel, whereas the second layer maps the feature map from the $2C$ channel back to the C channel. Channel expansion allows the LEA module to learn detailed channel relationships. Subsequently, the LEA module upsamples the feature maps ranked after the second convolutional layer to expand their spatial information and then uses the sigmoid function to map the learned convolutional relationships to the corresponding weights. Finally, the weights are loaded onto the original feature map by multiplication, which enhances information representation in the spatial and channel dimensions. The use of 3×3 convolutions in the LEA module allows the construction of local dependencies between adjacent parts on the feature map in the spatial dimension, as well as the construction of global relationships between channels in the channel dimension. In the semantic segmentation model, the weight construction of the spatial dependence of local and adjacent parts enhances the continuity of the object segmentation.

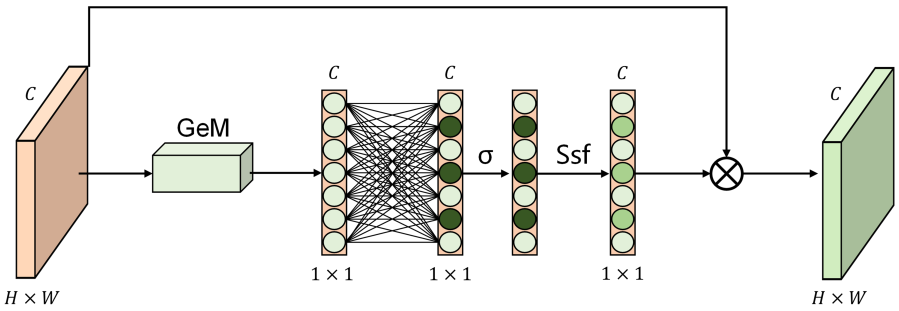

FIGURE 3. Non-salient feature extraction (NSF) module. "GeM" denotes generalized mean pooling. "σ" represents the Sigmoid function. "Ssf" is a function that suppresses salient features.

3.3. **Non-salient feature extraction (NSF) module.** The deep convolution operator extracts salient features of the target region. However, when a deep model encounters a target object with texture features similar to a non-target object, it often mistakenly recognizes the object as a non-target object. To improve the recognition ability of the semantic segmentation model for some floating rafts with similar features in the background, this study appropriately suppresses the salient features of floating rafts and uses non-salient features to explore the difficulty of identifying floating rafts. To achieve this, we built a non-salient feature (NSF) module, as shown in Figure 3. For the feature map $X \in \mathbb{R}^{H \times W \times C}$ input into the NSF module, where H, W and C represent the height, width and number of channels of the feature map, respectively, the NSF module first performs global generalized average pooling [48] on X along the spatial dimensions, compressing X into $Z \in \mathbb{R}^{1 \times 1 \times C}$. This process can be described as follows.

$$GeM = [(\frac{1}{X(n)} \sum_{x \in X(n)} x^{p(n)})^{\frac{1}{p(n)}}]_{n=1 \ldots C} \tag{3}$$

$$Z = GeM(X) \tag{4}$$

In equation (3), $X(n)$ represents the n-th channel of the feature map X. $P(n)$ represents the hyper-parameter corresponding to the n-th channel, which is a trainable parameter. For equation (3), when $p(n) = 1$, GeM performs global average pooling on feature map X, when $p(n) \to \infty$, GeM performs global max pooling on feature map X. The NSF module captures the most salient points on each channel of the feature map uniformly through GeM and uses this point to represent the corresponding channel feature. Next, the NSF module performs fully connected operations on Z to construct the relationship between each channel feature and all the other channel features of the feature map, thereby highlighting the salient features of the target. This process can be described by equation (5).

$$\omega_1 = \sigma(Dense(Z)) \tag{5}$$

$$\omega_2 = Ssf(\omega_1) \tag{6}$$

$$Ssf = \begin{cases} x \times \gamma, & x \geq mean(\omega_1) \\ x, & x < mean(\omega_1) \end{cases} \tag{7}$$

After the fully connected layer, Z is mapped to the corresponding weights ω_1 using a sigmoid function. In ω_1, a larger weight value indicates a more salient channel feature, enabling the localization of position information regarding the salient features of the target in the feature map X. Once the position of the salient features is determined, we use equations (6) and (7) to suppress them properly. Specifically, we first calculate the average weight of the weights on ω_1, then introduce a suppression factor γ, letting all weights greater than or equal to the average value be multiplied by γ, while keeping all weights smaller than the average value unchanged. This yields the suppressed weight ω_2 to suppress salient features in the feature map. Then, ω_2 is multiplied by the original feature map to suppress its salient features. Properly suppressing salient features can enable the model to extract more valuable target features.

3.4. **Binary cross-entropy loss function with boundary weight enhancement (BCE-BWE).** The boundary is the partition region between different classes, contains complex features and is susceptible to irregular shapes of the target. Pixels closer to the boundary were more likely to be incorrectly identified. These factors make it challenging for semantic segmentation models to accurately capture the complete contours of the object boundaries. This also applies to tasks involving the extraction of floating rafts from remote sensing images. The boundary features of floating rafts are severely affected by different objects in the background, resulting in incomplete extraction of floating raft boundaries. To enhance the ability of the model to identify floating raft boundaries, we generated boundary region weights from the true and predicted labels of the ocean remote sensing images. These weights guide the loss function for image segmentation, thereby providing stronger boundary supervision. Specifically, we first applied the Laplacian in a convolutional fashion to extract the boundary regions of the predicted true labels of ocean floating rafts. In this process, we extract the boundary information for different region sizes by adjusting the expansion rate of the Laplacian operator to obtain a complete representation of the boundary. The boundary extraction process is described by equation (8).

$$Lp = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{8}$$

$$Boundary = Conv(X, Lp)_d \tag{9}$$

Here, Lp represents the Laplacian operator, $Conv$ denotes a 2D convolution, X denotes the predicted or ground truth labels, and d represents the dilation rate of the convolution. The boundary map obtained through equation (9) is then converted into a binary mask, where 1 indicates the boundary and 0 represents the non-boundary regions. The boundary weight is then computed using the predicted boundary mask and the ground truth boundary mask, as described in equation (10):

$$W_{boundary} = \frac{\lambda_1 \times mask_p + \lambda_2 \times mask_g + 1}{\lambda_1 + \lambda_2 + 1} \tag{10}$$

$$L_{bce} = -(y \times \log(y_{pred}) + (1 - y) \times \log(1 - y_{pred})) \tag{11}$$

$$BCE - BWE = W_{boundary} \times L_{bce} \tag{12}$$

In equation (10), λ_1 and λ_2 are two hyperparameters that measure the difference between the boundary and non-boundary weights. $mask_p$ and $mask_g$ represent the binary masks of the predicted and ground truth boundary labels, respectively. After obtaining the boundary weight, we apply it to the binary cross-entropy loss function through multiplication, as shown in equations (11) and (12). The BCE-BWE term allows the boundary information to directly influence the loss function of the semantic segmentation model, enhancing the attention of the model towards building boundary. By improving the model's ability to recognize building boundary, BCE-BWE provides relatively less burden on the model by improving its ability to recognize building boundaries.

4. **Datasets and experiments.**

4.1. **Datasets.** The source satellite imagery for the WJD dataset was acquired by China's Gaofen-1 (GF-1) satellite. This study primarily utilized remote sensing images covering China's nearshore waters (with emphasis on coastal regions of Guangxi and Guangdong) along with collected ancillary background data as foundational inputs. Using ArcGIS and ENVI software, the WorldView-2 and WorldView-3 images underwent orthorectification, followed by data fusion of multispectral and high-resolution panchromatic bands via the PANSHARP method. Due to the large dimensions of the original remote sensing imagery, direct input into the network model for prediction would cause memory overflow. Consequently, the classification-ready images were first segmented into smaller tiles (processed in-memory without disk writing) matching the model's training sample size (e.g., 256×256 pixels). Each tile contains red, green, and blue (RGB) spectral bands at pixel-level resolution. The segmented tiles were subsequently annotated. The complete dataset comprises 10,930 images, partitioned into a training set (8,730 images), a validation set (1,100 images), and a test set (1,100 images). This dataset is proprietary and not publicly available.

4.2. **Evaluation Metrics.** Because we used an end-to-end pixel-level classification approach to extract floating rafts from remote sensing images, we selected pixel-level evaluation metrics for our experiments. The evaluation metrics in this study included precision, Recall, F1-score (F1), and Intersection over Union (IoU). These metrics are commonly used in the building extraction domains. Their definitions are as follows.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{15}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{16}$$

These include true prediction on a positive sample (TP), false prediction on a positive sample (FP), true prediction on a negative sample (TN), and false prediction on a negative sample (FN).

4.3. **Experimental details description.** All experiments in this study were conducted in Python 3.9, TensorFlow 2 and NVIDIA GeForce RTX 3090 (24G) environment. The initial learning rate was set as 0.001. When the training loss did not decrease for five epochs, the learning rate was reduced by half. We used the Adam optimizer and BCE-BWE in Section 3.4 to optimize the model training. The model was trained for 200 epochs, with a batch size of 6. For the hyperparameters in the model, γ in equation (7) and d in equation (9) were determined through ablation experiments. λ_1 and λ_2 in equation (10) were set to two, based on our experience.

5. Experimental Results and Discussion.

5.1. **Ablation Study.** The ablation study in this study was conducted using WJD datasets. All indicators in the experimental Table are expressed as percentages (%). The best results for each metric are highlighted in bold font in the Table. The ablation experiment consists of three parts. The first part discusses the selection of inhibitory factors that suppress the significant features in the NSF module. The second part focuses on the selection of size d for the central boundary region. The third section aims to demonstrate the influence of LEA, NSF, and BCE-BWE on the WNUNet model. The details are as follows.

(1) Comparing the effects of different suppression factors γ on the NSF module.

The experiments described in this section were conducted on the WNUNet model, using regular binary cross-entropy loss. By selecting different suppression factors γ, the impact of the NSF module on the WNUNet model was compared. The experimental results are listed in Table 1.

Table 1 presents the experimental results for the different inhibitory factors γ on the WJD dataset. In Table 1, "-" indicates NSF modules that do not suppress significant features, that is, the "Ssf" mapping in equation (5) is not applied. From Table 1 we can observe that compared with the NSF module without "Ssf" mapping, the IoU and F1 scores of the NSF module with "Ssf" mapping increased by more than 0.36% and 0.19% respectively. The improvements in these two indicators indicate that the utilization of

TABLE 1. Comparison of different γ on the WJD dataset.

Method	γ	IoU	F1	Recall	Precision
WNUNet	-	90.21	94.86	95.43	94.29
	0.0002	89.35	94.37	95.67	93.11
	0.002	89.71	94.57	95.25	93.90
	0.02	90.08	94.78	94.99	94.57
	0.2	90.16	94.82	94.82	93.64
	0.4	90.10	94.79	94.69	**94.89**
	0.6	90.54	95.03	95.86	94.22
	0.8	**90.57**	**95.05**	**96.42**	93.71

non-salient features can enhance the ability of the WNUNet model to construct recognition. However, the degree of suppression of significant features must be appropriately selected.

In Table 1, when $\gamma \leq 0.02$, the IoU and F1 metrics on the dataset decrease significantly. This is because over-suppressing the significant features of the raft will reduce the model's ability to learn the specific features of the raft, resulting in some backgrounds being incorrectly classified as rafts. However, when we appropriately suppressed the significant features, both the IoU and F1 metrics were improved on both datasets. For example, when $\gamma = 0.02$ or 0.8, the IoU, F1, and recall rate metrics on the WJD dataset improved. This indicates that by appropriately suppressing significant features, the model can learn more valuable features, particularly for floating rafts that are difficult to recognize by some models because they share similar texture features with the background. Regarding the selection of γ in the NSF module, we plotted the variations in the IoU metric under different γ values on the WJD dataset, as shown in Figure 4. However, when $\gamma = 0.8$, the WNUNet model achieved a significant performance improvement, with an IoU of 90.57% on the WJD dataset. However, the accuracy evaluation was only 93.71%, indicating a decrease. However, when the value of $\gamma = 0.6$, the WNUNet model showed relatively high recall rates and accuracy evaluation metrics on the dataset. By analyzing the IoU variation trend and accuracy rate on the dataset, we selected $\gamma = 0.6$ as the NSF module suppressor factor.

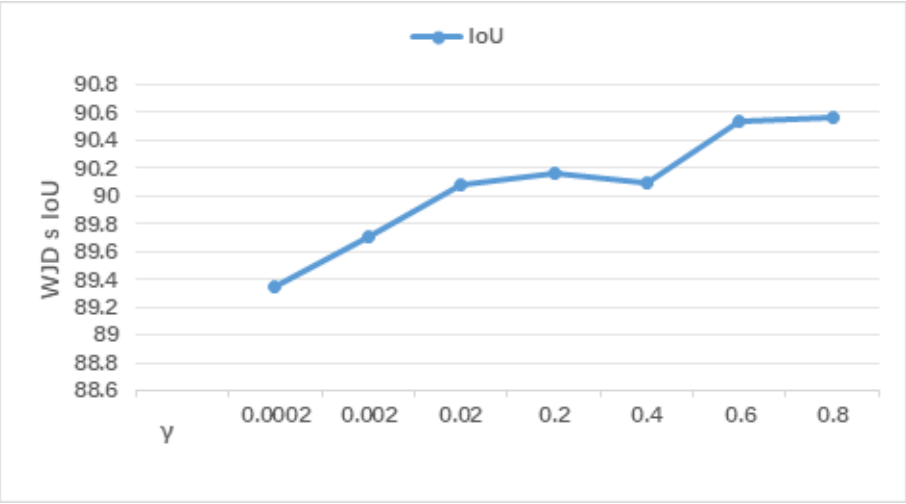

FIGURE 4. Variation of IoU(%) for different γ on WJD dataset.

(2) Comparing the effects of different boundary regions on the BCE-BWE.

The influence on the WNUNet model was observed by comparing the different boundary regions. Boundary regions of different sizes describe the boundary information of different degrees, and this information is affected by the expansion rate (d) in formula (9). Figure 5 shows the results of applying the Laplacian convolution with different d values to the labels in the WJD dataset. As shown in Figure 5, as d increased, the boundary area extracted from the label increased. Therefore, the enhancement

effect of the boundary weights on the binary cross-entropy loss also differs. The experimental results are presented in Table 2. "-" in Table 2 represents the cross-entropy loss without boundary enhancement.

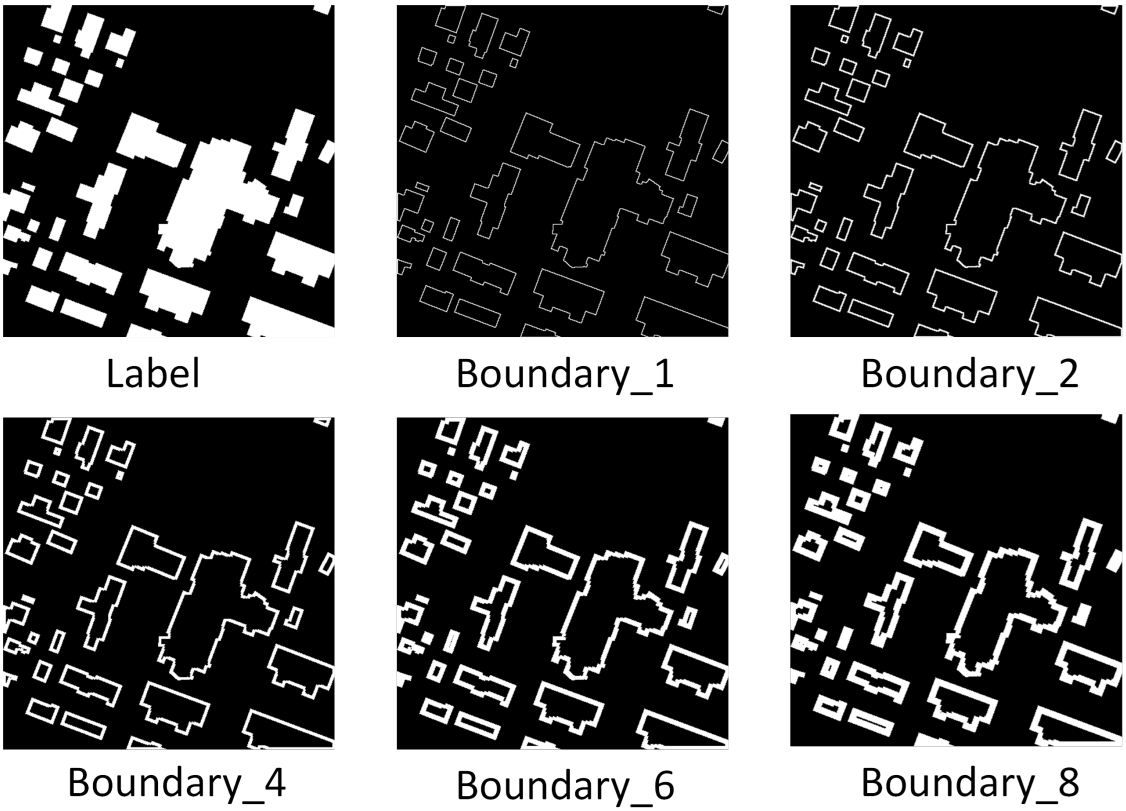

FIGURE 5. Boundary extraction of label using Laplacian convolution with different dilation rate. Boundary_1 and Boundary_2 correspond to the results with $d = 1$ and $d = 2$, respectively, and so on for other instances.

TABLE 2. Comparison of different d on the WJD dataset.

Method	***d***	**IoU**	**F1**	**Recall**	**Precision**
WNUNet + BCE – BWE	-	90.54	95.03	**95.86**	94.22
	1	90.59	95.06	95.54	94.59
	2	90.72	95.13	95.13	95.13
	4	**90.91**	**95.24**	95.70	94.79
	6	90.71	95.13	94.78	95.48
	8	90.71	95.13	94.77	**95.49**

As shown in Table 2, compared with the absence of boundary weight enhancement, the use of boundary weight enhancement improved the IoU and F1 metrics of the WNUNet model on the WJD dataset. In the WJD dataset, when $d = 4$ occurred, the improvements in the IoU and F1 indicators were the most significant, increasing by 0.37% and 0.21% respectively. This demonstrates the effectiveness of the proposed method. Furthermore, the IoU and F1 indicators in Table 2 improved further with an increase in d. For example, when $d = 1, 2$ and 4. This indicates that expanding the boundary region can enhance the boundary weights. However, when the scale was too large, the IoU and F1 indicators declined. For example, on the WJD dataset, when $d > 4$, IoU and F1 decrease. This is because the value of d is too large, resulting in a large acceptance domain for the Laplacian convolution. When extracting the boundaries of the small floating raft, it is completely extracted and is not affected by the boundary weights. For example, Figure 5 shows some examples of small floating rafts in Boundary_6 and

Boundary_8. According to Table 2, the experimental results for $d = 4$ were selected for this study. In this case, WNUNet achieved IoU and F1 metrics of 90.91% and 95.24% respectively on the WJD dataset, with relatively good experimental indicators.

(3) Ablation study of the LEA module, NSF module, and BCE-BWE on the WNUNet model.

To demonstrate the effectiveness of the proposed LEA and NSF modules, we conducted an ablation study by gradually adding each module to the baseline model to clarify their respective contributions. The experimental results are listed in Table 3. The baseline in WNUNet has a structure similar to that of the UNet. The key difference is that each layer of the baseline encoder is composed of three 3×3 convolutional layers. It can be observed from Table 3 that gradually incorporating LEA, NSF, and BCE-BWE modules into the baseline model will lead to an improvement of the IoU and F1 metrics on the WJD dataset. Furthermore, adding the LEA and NSF modules increased the traffic and parameters of the baseline model. However, overall, WNUNet's FLOPs and parameters are relatively low.

TABLE 3. Ablation study for each module.

Method	Dataset	LEA	NSF	BCE-BWE	IoU	F1	FLOPs(G)	Params(M)
Baseline	WJD				89.32	94.36	131.08	8.64
+LEA	WJD	√			89.99	94.78	150.44	11.78
+LEA+NSF	WJD	√	√		90.54	95.03	150.50	11.88
+All (WNUNet)	WJD	√	√	√	**90.91**	**95.24**	150.50	11.88

After adding the LEA module, IoU and F1 of the baseline model on the WJD dataset increased by 0.67% and 0.42% respectively. This demonstrates the effectiveness of the LEA module. The LEA module focuses on the information channels and spatial dimensions and is used to aggregate the local information feature map and construct the relationship between the local and adjacent local areas, enabling the LEA module to enhance the semantic segmentation model for the continuous extraction of the floating raft. In examples I and II of Figure 6, compared with the predictions without the LEA module in column c, the continuity of some rafts in columns d-f in the predictions with the LEA module was improved. This indicates that the LEA module can enhance the ability of the baseline model to recognize the continuity of the floating raft.

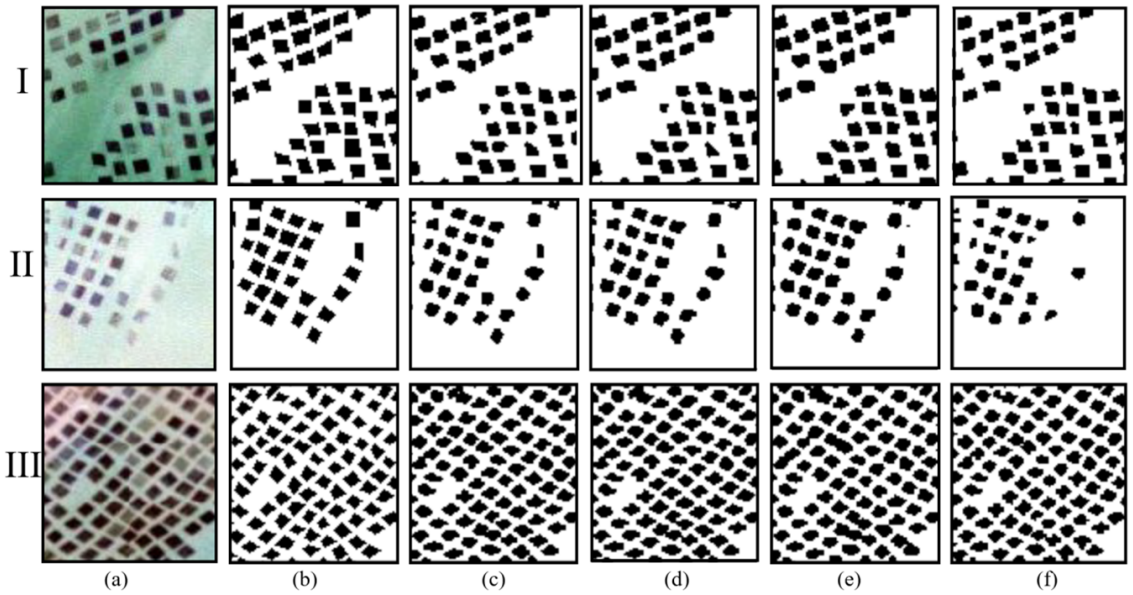

FIGURE 6. Prediction plots of ablation studies for each module on the WJD dataset. (a) Original image. (b) Ground Truth. (c) Baseline. (d) Baseline+LEA. (e) Baseline+LEA+NSF. (f) Baseline+LEA+NSF+ BCE-BWE.

Compared with the baseline using only the LEA module, the baseline using the additional NSF module increased IoU and F1 on the WJD dataset by 0.55% and 0.25% respectively. This proved the validity of the NSF module in the WNUNet model. The NSF module effectively suppressed the significant features of the floating raft and utilized non-significant features to enable the model to learn more valuable discrimination features. This enhances the ability of the model to identify raft areas with similar background features. The baseline without the NSF module misclassifies some floating rafts in some areas as the background, where the baseline with the NSF module shows a better recognition ability for this area.

The boundary weights are obtained by comparing the predicted boundary images with the real boundary images, and then the boundary weights are used to enhance the binary cross-entropy loss, thereby increasing the focus of the model on constructing boundaries. As shown in Table 3, compared to the baselines using only the LEA and NSF modules, the IoU of the added baselines on the WJD dataset increased by 0.37% and 0.21% respectively. Furthermore, the addition of BCE-BWE did not increase the FLOPs or number of parameters. The enhancement effect on the construction boundary is illustrated in Example III in Figure 6.

5.2. **Comparison experiment.** Through the ablation. We determined the suppression factor of the LEA module and the dilation rate d in BCE-BWE, thus clarifying the structure of the WNUNet model. To further validate the effectiveness of WNUNet, we compared it with state-of-the-art and classical building extraction methods on the WHU and Inria datasets. We proved the effectiveness of WNUNet by comparing various evaluation metrics and prediction results of different methods on the established dataset. The experimental results on the WJD dataset are shown in Table 4 as follows.

Compared with DeeplabV3+, the classic raft extraction models UNet and HRNetV2 demonstrated relatively outstanding performance. This highlights the advantages of UNet and HRNetV2 for multi-scale skip connections in the field of raft extraction. Therefore, we tested the feasibility of the WNUNet model by embedding modules, such as LEA and NSF, into the construction of the Unet architecture to construct the extraction task. Compared with classical methods, methods such as B-FGC-Net, Swin-U-S, MAP-Net, MSRF-Net, and WNUNet have already included attention mechanisms that focus on the channel dimension, spatial dimension or global context information. Therefore, compared with classical methods, these methods have significantly improved the establishment of evaluation metrics, such as IoU and F1 scores. Our WNUNet model includes an LEA module, which considers both local and global channel information simultaneously. Compared with attention modules in models such as B-FGC-Net, MAP-Net and MSRF-Net, which only focus on single-dimensional information in channels or Spaces, the LEA module shows better performance. However, the difference between our WNUNet and other raft extraction methods lies in the use of non-significant features. Traditional raft extraction algorithms typically employ various techniques to enable the model to learn more significant raft features. In contrast, our WNUNet model uses the NSF module to appropriately suppress the significant features of the raft, thereby the model to obtain more distinguishable raft features. Although methods such as Swin-U-S focus on the global context information of feature mapping, they ignore the importance of non-significant features. Therefore, our WNUNet method achieved good results in terms of the IoU, F1, and recall metrics on the WJD construction dataset. The MAPNet method retains accurate spatial information in feature mapping through a multi-branch parallel structure and achieves good results. However, when dealing with large-scale remote sensing images, The complexity of the multi-branch parallel structure is much higher than that of structures such as UNet.

Table 4. Comparison of different methods on the WJD dataset.

Method	IoU	F1	Recall	Precision
UNet	89.05	94.21	94.32	94.10
HRNetV2	89.10	94.24	95.55	92.96
DeeplabV3+	87.81	93.51	93.67	93.36
B-FGC-Net	90.04	94.76	94.49	95.03
Swin-U-S	90.68	95.11	95.18	95.04
MAP-Net	90.86	95.21	94.81	**95.62**
MSRF-Net	90.16	94.82	94.68	94.97
WNUNet (Ours)	**90.91**	**95.24**	**95.70**	94.79

The predicted maps of WNUNet on the WJD dataset in comparison with experimental results are shown in Figure 7. The design purpose of WNUNet is to improve the recognition ability of the classical semantic segmentation model, which is used to construct continuity, construct boundaries and floating raft parts with similar features to the background. For example, in example I in Figure 7, UNet, HRNetV2, and DeeplabV3+ show discontinuous recognition of individual rafts, whereas the predicted maps of WNUNet alleviate this issue significantly owing to the LEA module. The recognition of floating rafts with textural features similar to the background is also a problem. For example, some rafts have very similar textures to the background. This led to the classic semantic segmentation model incorrectly identifying this part of the raft. However, our WNUNet model uses the NSF module to identify rafts similar to the background. The NSF module is helpful for extracting constructed non-significant features, enabling the model to learn more valuable features. To identify the floating raft boundary, WNUNet utilizes this, which enables the model to segment a more complete boundary contour. Examples II and III are shown in Figure 7.

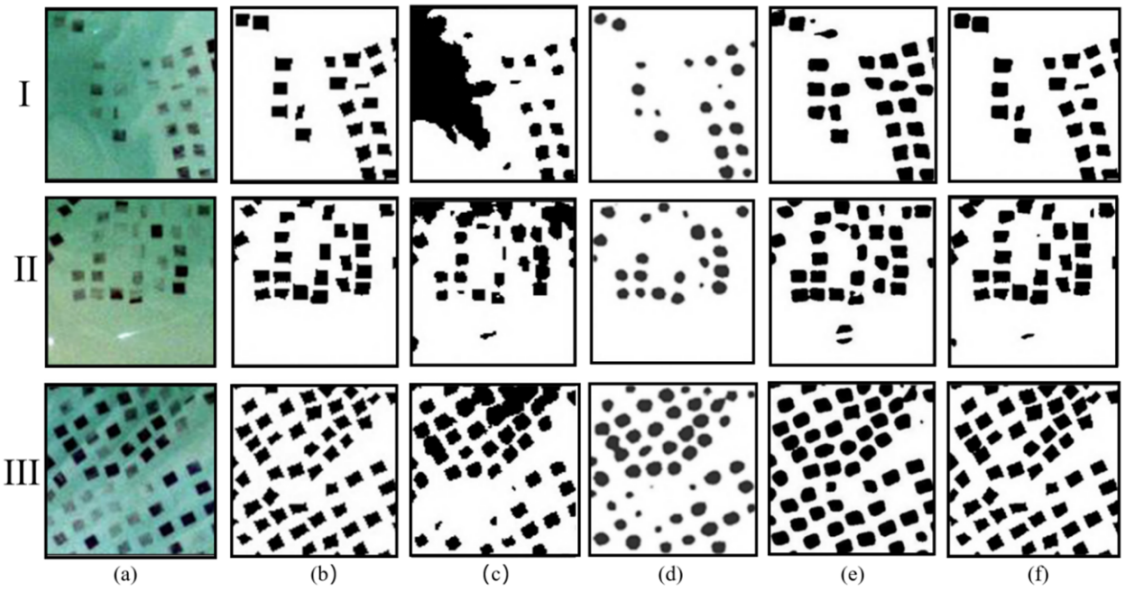

FIGURE 7. Prediction plots of different methods on the WHU dataset. (a) Original image. (b) Ground Truth. (c) UNet. (d) HRNetV2. (e) DeepLabV3+. (f) WNUNet.

6. **Conclusions.** This study proposes three functional modules, LEA, NSF, and BCE-BWE, to address the challenge of extracting floating rafts and integrating them into the UNet architecture to create an efficient WNUNet model. The LEA module was designed to enhance the model's ability to recognize the continuity of floating rafts. By establishing the dependency relationship between adjacent local spatial information and global channel information, the LEA module enhances the correlation among the internal regions of the raft and ensures coherence of the segmentation. The NSF module utilizes inhibitory factors to suppress the significant features of the raft and highlights its non-significant features, enabling WNUNet to learn more valuable characteristics of the raft. This plays an important role in extracting floating rafts with textural features similar to those of the background. Binary cross-entropy loss with enhanced boundary weights was adopted, and the boundary information of the predicted map is directly incorporated into the segmentation loss. This increases the focus of the model on the boundary of the floating raft and improves its ability to recognize the boundary of the floating raft. Through an ablation study of the WJD dataset, we proved the effectiveness of each module in WNUNet. WNUNet integrates the advantages of the LEA, NSF, and UNet modules, which enables the experimental results on the WJD dataset to be superior to those of recent mainstream methods and classic raft extraction methods. However, achieving high-performance segmentation on the raft dataset still requires a large amount of data and corresponding annotations. Therefore, introducing unsupervised or semi-supervised methods into WNUNet will be a very good research direction for the future.

7. **Acknowledgment.** This research was supported by the Key Research Project of Natural Science of Universities in Anhui Province (2024AH050607), the Training Program for Academic (Professional) Leaders in Anhui Province (DTR2025060), and the Scientific Research Innovation Team for Intelligent Information Processing in Anhui Province (2024AH010012).

REFERENCES

[1] Yu, Jintao, et al. "Nursing Effects of Large-Scale Floating Raft Aquaculture Habitats on Conger myriaster: A Perspective from Marine Ranching." *Water,* 17.3: 431, 2025.

[2] Yu, Haomiao, et al. "MSSFNet: A multiscale spatial–spectral fusion network for extracting offshore floating raft aquaculture areas in multispectral remote sensing images." *Sensors,* 24.16: 5220, 2024.

[3] Lozano, Jose, Diego Renza, and Dora M. Ballesteros L. "Evaluation of pruning according to the complexity of the image dataset." *J. Inf. Hiding Multim. Signal Process.,* 15.2 (2024): 53-62, 2024.

[4] Zou Guohua, Zhang Shuoyuan, Wang Fei & Zhang Shengmao. "An empirical study of Kelp raft farming area monitored by satellite remote sensing". *China Fisheries,* (05), 73-77, 2022.

[5] Chu Jialan, Shao Guanghui, Zhao Jianhua, Gao Ning, Wang Fei & Cui Bingge. "Information Extraction Method for Floating Raft Aquaculture of Gaofen No.1." *Surveying and mapping science,* 45 (01), 92-98.

[6] Wang, Kun, et al. "Application of a VOF multiphase flow model for issues concerning floating raft aquaculture." *Water,* 15.19: 3450, 2023.

[7] Xian Wang, Yanan Wang, Zhengzhong Li, Hailong Che, Lanlan Zhu, Tongfei Sheng, Hua Zhou, Duanyang Geng. "Research on the three-body kelp harvesting ship based on floating raft aquaculture mode". *Aquacultural Engineering,* 110, 102545-102545, 2025.

[8] Jiahua Wan, and Vladimir Y. Mariano. "Monitoring of floating structures using segmentation and object classification of remotely-sensed images." *Journal of Physics: Conference Series,* 2858.1: 012008-012008, 2024.

[9] Chen O. "Information extraction of offshore aquaculture areas based on coupling of multi-features and object classification". *Shanghai ocean university,* 2023.

[10] X. Wang, M. Kang, Y. Chen, W. Jiang, M. Wang, T. Weise et al., "Adaptive Local Cross-Channel Vector Pooling Attention Module for Semantic Segmentation of Remote Sensing Imagery", *Remote Sens.,* vol. 15, no. 8, Apr, 2023.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 3431-3440, 2015.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent,* Nov. pp. 234-241, 2015.

[13] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *in Proc. Eur. Conf. Comput. Vis. (ECCV),* pp. 801-818, 2018.

[14] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 2117-2125, 2017.

[15] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* Jun, pp. 5693-5703, 2019.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al., "Attention Is All You Need," *in Proc. Adv. Neural Inf. Process. Syst.,* Long Beach, CA, Dec, pp. 5998–6008, 2017.

[17] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *in 35th Conference on Neural Information Processing Systems (NeurIPS),* Electr Network, Dec, 2021.

[18] L. Wang, R. Li, C. Zhang, H. Fang, C. Duan, X. Meng et al., "UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS J. Photogramm. Remote Sens.,* vol. 190, pp. 196-214, 2022.

[19] H. Yan, C. Zhang, and M. Wu, "Lawin transformer: Improving semantic segmentation transformer with multi-scale representations via large window attention," *arXiv:2201.01615,* 2022.

[20] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang et al., "Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* Electr Network, Jun, pp. 6877-6886, 2021.

[21] M. Guo, T. Xu, J. Liu, Z. Liu, P. Jiang, T. Mu et al., "Attention mechanisms in computer vision: A survey," *Computational Visual Media,* vol. 8, no. 3, pp. 331-368, 2022.

[22] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang et al., "Dual Attention Network for Scene Segmentation," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* Long Beach, CA, Jun, pp. 3141-3149, 2019.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* 2016, pp. 770-778.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 7132-7141, 2018.

[25] X. Cheng, X. Li, J. Yang, and Y. Tai, "SESR: Single image super resolution with recursive squeeze and excitation networks," *in 24th International conference on pattern recognition (ICPR),* pp. 147-152, 2018.

[26] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 11534-11542, 2020.

[27] S. H. Woo, J. C. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," *in Proc. Eur. Conf. Comput. Vis. (ECCV),* pp. 3-19, 2018.

[28] Q. Hou, D. Zhou, and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* Electr Network, pp. 13708-13717, 2021.

[29] Z. Zhong, Z. Lin, R. Bidart, X. Hu, I. Daya, Z. Li et al., "Squeeze-and-attention networks for semantic segmentation," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR),* pp. 13065-13074, 2020.

[30] Q. Zhu, C. Liao, H. Hu, X. Mei, and H. Li, "MAP-Net: Multiple Attending Path Neural Network for Building Footprint Extraction From Remote Sensed Imagery," *IEEE Geosci. Remote Sens.,* vol. 59, no. 7, pp. 6169-6181, 2021.

[31] Y. Zhao, G. Sun, L. Zhang, A. Zhang, X. Jia, and Z. Han, "MSRF-Net: Multiscale Receptive Field Network for Building Detection From Remote Sensing Images," *IEEE Geosci. Remote Sens.,* vol. 61, 2023.

[32] Y. Wang, X. Zeng, X. Liao, and D. Zhuang, "B-FGC-Net: A Building Extraction Network from High Resolution Remote Sensing Imagery," *Remote Sens.,* vol. 14, no. 2, Jan, 2022.

[33] Y. Qiu, F. Wu, J. Yin, C. Liu, X. Gong, and A. Wang, "MSL-Net: An Efficient Network for Building Extraction from Aerial Imagery," *Remote Sens.,* vol. 14, no. 16, Aug. 2022.

[34] Y. Qiu, F. Wu, H. Qian, R. Zhai, X. Gong, J. Yin et al., "AFL-Net: Attentional Feature Learning Network for Building Extraction from Remote Sensing Images," *Remote Sens.,* vol. 15, no. 1, Jan, 2023.

[35] J. Cai, and Y. Chen, "MHA-Net: Multipath Hybrid Attention Network for Building Footprint Extraction From High-Resolution Remote Sensing Imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.,* vol. 14, pp. 5807-5817, 2021.

[36] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv:2010.11929,* 2020.

[37] L. Wang, S. Fang, X. Meng, and R. Li, "Building Extraction With Vision Transformer," *IEEE Geosci. Remote Sens.,* vol. 60, 2022.

[38] L. Xia, S. Mi, J. Zhang, J. Luo, Z. Shen, and Y. Cheng, "Dual-Stream Feature Extraction Network Based on CNN and Transformer for Building Extraction," *Remote Sens.,* vol. 15, no. 10, May, 2023.

[39] C. Qiu, H. Li, W. Guo, X. Chen, A. Yu, X. Tong et al., "Transferring Transformer-Based Models for Cross-Area Building Extraction From Remote Sensing Images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.,* vol. 15, pp. 4104-4116, 2022.

[40] Fei Fu, Xiaoying Zhang, Zhe Hu, Yan Li, Lihe Wang & Jianxing Yu. "Research on Sea Trial Techniques for Motion Responses of HDPE Floating Rafts Used in Aquaculture". *Journal of Marine Science and Engineering,* 12(7), 1150-1150, 2024.

[41] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 12, pp. 2481-2495, Dec, 2017.

[42] B. Ma, and C. Chang, "Semantic Segmentation of High-Resolution Remote Sensing Images Using Multiscale Skip Connection Network," *Ieee Sensors Journal,* vol. 22, no. 4, pp. 3745-3755, 2022.

[43] R. Li, C. Duan, S. Zheng, C. Zhang, and P. M. Atkinson, "MACU-Net for Semantic Segmentation of Fine-Resolution Remotely Sensed Images," *IEEE Geosci. Remote Sens. Lett.,* vol. 19, 2022.

[44] F. Chen, H. Liu, Z. Zeng, X. Zhou, and X. Tan, "BES-Net: Boundary Enhancing Semantic Context Network for High-Resolution Image Semantic Segmentation," *Remote Sens.,* vol. 14, no. 7, 2022.
[45] J. Hoin, C. Han-Soo, and K. Myung-joo, "Boundary Enhancement Semantic Segmentation for Building Extraction From Remote Sensed Image," *IEEE Geosci. Remote Sens.,* vol. 60, pp. 1-12, 2021.
[46] S. He, and W. Jiang, "Boundary-assisted learning for building extraction from optical remote sensing imagery," *Remote Sens.,* vol. 13, no. 4, pp. 760, 2021.
[47] A. Li, L. Jiao, H. Zhu, L. Li, and F. Liu, "Multitask semantic boundary awareness network for remote sensing image segmentation," *IEEE Geosci. Remote Sens.,* vol. 60, pp. 1-14, 2021.
[48] M. Berman, H. Jégou, A. Vedaldi, L. Kokkinos, and M. Douze, "Multigrain: a unified image embedding for classes and instances," *arXiv:1902.05509,* 2019.