TRANSNAV http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 19 Number 2 June 2025
DOI: 10.12716/1001.19.02.30

Edge-Guided Multi-Scale Fusion and Importance-Aware Learning for Real-Time Semantic Segmentation in Waterborne Navigation

L. Chen[1], J. Zou[1], Y. Huang[2], Y. Zhou[1], G. Hao[1] & Y. Zhang[3]
[1] *Wuhan University of Technology, Wuhan, China*
[2] *Wuhan Institute of Shipbuilding Technology, Wuhan, China*
[3] *Hubei University of Chinese Medicine, Wuhan, China*

ABSTRACT: Effective multi-scale feature representation and focused attention on critical objects are essential for accurate perception of waterborne navigation scenes. To address the insufficient exploitation of multi-scale information in existing methods that leads to imprecise segmentation, this study proposes a real-time semantic segmentation method for waterborne navigation scenes through multi-scale information enhancement and importance-weighted optimization. First, DDRNet-23-slim is selected as the backbone network for feature extraction. An edge-guided branch is embedded into its shallow layers, and a Dynamic Feature Fusion Module (DFFM) is constructed by integrating a lightweight hybrid attention mechanism, effectively enhancing multi-scale feature interaction capabilities. Second, the loss function is improved using an importance-weighted strategy to prioritize critical objects during training. Finally, a parameter-free attention mechanism is introduced in the upsampling stage, maintaining real-time performance while ensuring segmentation stability for key objects under complex background interference. Evaluations on the On_Water and Seaships datasets demonstrate that the proposed method achieves mIoU scores of 83.1% and 73.2%, respectively, with ship segmentation accuracy reaching 88.2% on On_Water. The inference speed attains 69.1 FPS, outperforming mainstream real-time segmentation models (e.g., DDRNet, STDC) in balancing accuracy and efficiency. Notably, it exhibits stronger robustness in complex inland river scenarios with dense shore structures and numerous small targets.

1 INTRODUCTION

With the rapid development of artificial intelligence and automation technologies, intelligent shipping and unmanned vessels are gradually moving toward practical applications, bringing new opportunities to fields such as water traffic management, autonomous ship navigation, and ocean monitoring [1]. Achieving autonomous navigation and obstacle avoidance for unmanned vessels relies on the precise understanding of various objects in maritime traffic scenes, such as ships, buoys, water bodies, and the sky. Semantic segmentation technology, as a key means to achieve this goal, is gaining widespread attention in maritime scenarios [2].

Early studies employed traditional digital image processing methods for water body detection and obstacle recognition. For example, P. Santana et al. [3] utilized digital image processing techniques for water body detection, and Fefilatye et al. [4] proposed obstacle detection methods based on handcrafted features. However, these methods' reliance on simple features leads to significant declines in segmentation accuracy in complex scenarios such as coastal areas, near-shore regions, or docks, especially under conditions of visual ambiguity and light reflections. Additionally, with the widespread adoption of high-resolution devices, traditional methods face challenges of slow processing speeds and poor segmentation performance when handling large-scale image data.

In recent years, semantic segmentation methods based on deep convolutional neural networks (CNNs) have achieved remarkable results in terrestrial traffic scenes due to their powerful feature learning capabilities. However, directly applying these methods (e.g., PSPNet) to maritime traffic scenes still presents multiple challenges. Cheng et al. [5] proposed a deep network-based method for land-sea boundary segmentation in high-resolution remote sensing images, but it often suffers from underfitting and insufficient dynamic water surface feature extraction in maritime scenes. C.Y. Jeong et al. [6] used the Pyramid Scene Parsing Network (PSPNet) for horizon detection, which improved boundary recognition to some extent, but misjudgments still occur under conditions of visual ambiguity or unclear boundaries between the horizon and the sky. Meanwhile, M. Kristan et al. [7] monocular obstacle detection model struggles with large sea surface fluctuations and small obstacle recognition. Some studies have attempted to enhance segmentation performance by incorporating multimodal data. For instance, Scherer et al. [8] fused data from stereo cameras, IMU/GPS, and laser scanners, achieving certain improvements. However, these methods heavily rely on external hardware and complex post-processing steps, making real-time segmentation difficult in resource-constrained environments. Borja Bovcon et al. [9] improved segmentation accuracy under visually ambiguous conditions by designing a semantic separation loss function combined with high-precision IMU data, but such reliance on external high-precision data limits their application on small devices like mobile platforms. Furthermore, to address issues such as water surface reflections and boundary ambiguity, Qiao Yulong et al. [10] improved water surface segmentation accuracy through image preprocessing and sea-sky line estimation parameters. Bao Xuecai et al. [11] introduced attention mechanisms and fully connected conditional random field models to enhance floating object boundary recognition, while Xiong Rui et al. [12] designed a fast feature extraction network to balance real-time performance and accuracy. However, these methods still fall short in effectively fusing multi-scale information, focusing on key targets, and ensuring overall system real-time performance, making it difficult to fully meet the dual requirements of segmentation accuracy and real-time performance for autonomous navigation and obstacle avoidance of unmanned vessels.

In summary, although some solutions for semantic segmentation in maritime navigation scenes have been proposed, existing methods still struggle to achieve ideal segmentation results due to challenges such as complex lighting variations, adverse weather conditions, dynamic backgrounds, and small target detection in maritime environments. Some scholars have directly borrowed segmentation methods from terrestrial traffic scenes, but these methods fail to adequately account for interference factors such as water surface reflections and rain fog in maritime environments, leading to issues like edge blurring and inaccurate recognition in practical applications. Additionally, given the limited computational resources of small unmanned vessels and the demand for real-time processing, there is an urgent need to develop an efficient and specialized real-time semantic segmentation method for maritime navigation scenes. To this end, this paper proposes a real-time semantic segmentation model for maritime navigation scenes that integrates multi-scale information and importance weighting (Multi-Scale Importance-Aware Network, MSIA-Net). The model aims to address edge blurring in object segmentation, enhance recognition accuracy for complex shore backgrounds and multi-scale object contours, and thereby provide more accurate and real-time environmental understanding for autonomous navigation and obstacle avoidance of unmanned vessels.

2 MODEL NETWORK ARCHITECTURE DESIGN

To address the challenges of numerous floating objects, frequent small obstacles, and complex background interference in maritime navigation scenes while ensuring real-time segmentation performance, this paper introduces an edge-guided branch and a dynamic feature fusion module into the backbone network of DDRNet-23-slim, achieving precise capture of target boundaries and multi-scale information. Specifically, the edge-guided branch employs the Sobel operator for edge detection and integrates the SE module for channel attention adjustment, generating edge-aware feature maps to enhance the model's ability to detect small obstacles and floating objects. The dynamic feature fusion module combines the edge-guided module with multi-scale features from the backbone network, adaptively generating weights for features at different scales through a lightweight attention network, thereby achieving efficient fusion and further improving the model's segmentation performance for multi-scale objects.

Additionally, to enhance the quality of feature maps without increasing model complexity, a parameter-free attention mechanism is introduced during the upsampling phase. This mechanism adjusts feature maps in both channel and spatial dimensions, effectively enhancing their representational capacity while maintaining the model's lightweight design.

Finally, to further improve the model's performance in complex background and multi-scale object segmentation tasks, an improved loss function based on importance weighting is proposed. This loss function assigns different weights to different pixels, enabling the model to focus more on pixels that significantly impact segmentation results during training, thereby enhancing the model's robustness and accuracy. The overall architecture of MSIA-Net is illustrated in Figure 1.

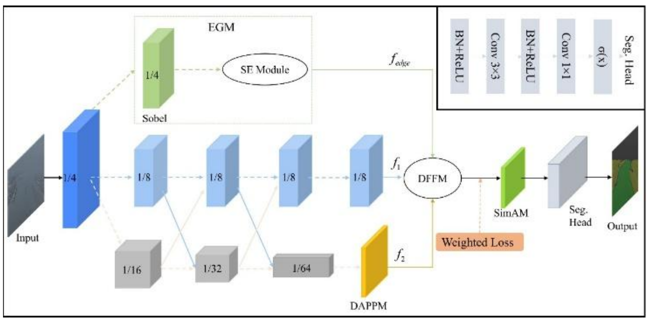

Figure 1. Network structure of MSIA

2.1 *Enhancement of the Backbone Network with Multi-Scale Information*

To adapt to the segmentation tasks of ships of varying sizes and complex shore backgrounds in maritime navigation scenes, this paper improves the backbone network of DDRNet by introducing an edge-guided branch and integrating multi-scale information through a dynamic feature fusion module, thereby enhancing the model's segmentation accuracy for ships, obstacles, and shore boundaries in complex water scenes. Specifically, by incorporating the Edge Guidance Module (EGM), the model's sensitivity to water surfaces, waves, and small object boundaries is significantly enhanced. Combined with the dynamic feature fusion module, multi-scale contextual information is dynamically aggregated, expanding the receptive field and improving the model's ability to model complex background regions.

2.1.1 *Edge-guided Branch*

To address the segmentation task of ships of varying sizes and complex shoreline backgrounds in waterborne navigation scenarios, this paper improves the backbone network of DDRNet-23-slim by introducing an edge-guided branch (whose structure is shown in Figure 2). The EGM is the core component of the edge enhancement branch, aiming to explicitly enhance edge information to improve the model's segmentation accuracy of target boundaries in complex scenes. This module first converts the three-channel RGB image into a grayscale image, then uses the Sobel operator to perform edge detection on the grayscale image to enhance the edge response. Subsequently, the Squeeze-and-Excitation (SE) module is employed to adaptively weight the edge features, ensuring that edge information plays a more significant role in the subsequent feature fusion process.

The Sobel operator is a widely used edge detection method in the field of image processing. It primarily relies on discrete differential operators to approximate the first-order gradient magnitude and direction at each pixel location in an image. The Sobel operator employs a pair of 3×3 convolution kernels to estimate the gradient magnitudes in the horizontal directions and vertical directions , respectively:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * f_{input} \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * f_{input} \tag{1}$$

where * denotes the convolution operation, and f_{input} represents the input feature map. These two convolution kernels highlight regions with rapid intensity changes along specific directions, thereby identifying edges in the image. To ensure that edge information plays a more significant role in the subsequent feature fusion process, the EGM module adaptively weights the edge features using the Squeeze-and-Excitation (SE) module. The SE module computes channel-wise weights by performing global average pooling on the feature map, followed by weighted fusion of the features to enhance the important characteristics. he computational process of this module is as follows:

$$f_{edge} = SE\left(Sobel(f_{\frac{1}{4}})\right) \tag{2}$$

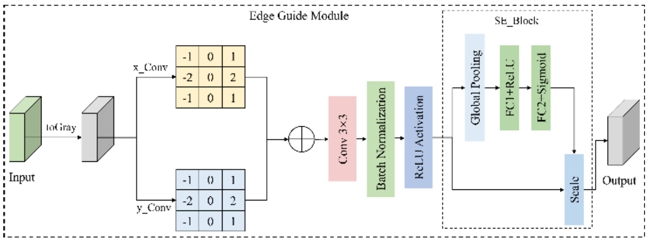

Figure 2. Edge-guided Branch

2.1.2 *Dynamic Feature Fusion Module with Multi-scale Information Enhancement*

In the context of semantic segmentation for waterborne navigation scenes, the fusion of multi-scale information presents a significant challenge. Waterborne environments typically encompass a variety of features at different scales, such as intricate shoreline structures, water surfaces, and vessels, each carrying distinct semantic information across scales. To address this, we propose an enhanced dynamic feature fusion module (whose structure is shown in Figure 3) designed to dynamically integrate features from multiple levels. This module assigns appropriate fusion weights to different spatial locations, effectively capturing positional discrepancies between feature maps and the original input image.

The implementation begins with upsampling feature maps of different scales, f_1 and f_2 , to the size of f_{edge} followed by concatenation along the channel dimension. A 1×1×1 spatial convolution kernel is then applied to the concatenated feature map, and a Sigmoid function is used to compute the spatial weight w_{sp} which enhances the representation of critical spatial regions. The spatial weight w_{sp} is calculated as:

$$w_{sp} = \delta\left(Conv_{1\times1\times1}\left(concat(f_1; f_2; f_{edge})\right)\right) \tag{3}$$

where δ denotes the Sigmoid function. Concurrently, channel weights w_{ch} are derived through a cascade of average pooling (AVGPool), a convolutional layer (Conv), and a Sigmoid activation. The channel weight w_{ch} is expressed as:

$$w_{ch} = \delta\left(Conv_{1\times1}\left(AVGPool\left(concat(f_1; f_2; f_{edge})\right)\right)\right) \tag{4}$$

Subsequently, each feature map is weighted using the computed spatial weights w_{sp} and channel weights w_{ch}, followed by element-wise addition. A 1×1 convolution is applied to the fused feature map to compress its channels, yielding the final feature map F_{end}. The final output is computed as:

$$f_{end} = Conv_{1\times1}\left((f_1 \oplus f_2 \oplus f_{edge}) \otimes w_{sp} \otimes w_{ch}\right) \tag{5}$$

By generating dynamic weights, the proposed module adaptively adjusts the importance of each feature based on the global context of the input. This enables the model to selectively emphasize key features, thereby enhancing its ability to represent multi-scale information effectively. The dynamic feature fusion mechanism ensures robust performance in capturing diverse and complex features inherent in waterborne navigation scenes.

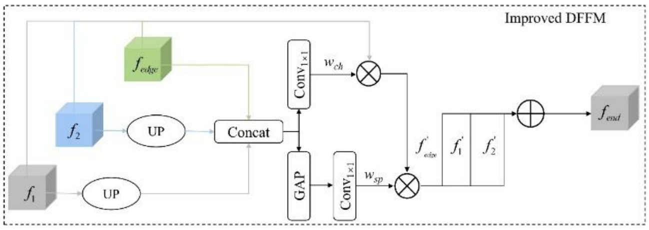

Figure 3. Improved Dynamic Feature Fusion Module

2.2 *Parameter-free Attention Guidance*

In real-time semantic segmentation for waterborne navigation scenarios, key targets (such as ships, obstacles, and navigation marks) typically occupy only a small portion of the image, while the majority of the area consists of water surfaces. This not only results in relatively sparse information about key targets in the overall image but also introduces interference in feature extraction due to factors such as water reflections, waves, and environmental lighting changes. Traditional attention mechanisms (e.g., spatial attention or channel attention) often rely on additional convolutional or fully connected layers to generate attention weights, thereby increasing model parameters and computational overhead, which is not ideal for real-time segmentation tasks.

To address this issue, the SimAM (Simple Parameter-Free Attention Module) is introduced. It constructs an energy function to measure the importance of each neuron(whose structure is shown in Figure 4), enabling direct attention weighting on the original three-channel feature maps without introducing additional parameters. The core idea of SimAM is to evaluate the importance of a neuron based on its ability to distinguish itself from other neurons within the same channel (its structure is shown in the figure). Specifically, if a neuron can be clearly distinguished from other neurons in the same channel (i.e., it exhibits strong linear separability), it indicates that the neuron is more active in expressing key features and its information is more representative. Conversely, neurons with lower discriminability are considered less important. By dynamically assessing the importance of each neuron, SimAM can perform fine-grained spatial and channel-wise weighting adjustments on the original feature maps without adding extra parameters. This enhances the focus on key targets while suppressing irrelevant background information. The specific implementation is as follows:

$$E(t)=\min_{w,b}\left\{\left(w\cdot t+b-1\right)^2+\sum_{i=1}^{N-1}\left(w\cdot o_i+b+1\right)^2+\lambda w^2\right\} \tag{6}$$

where N represents the number of neurons, t represents the output of the target neuron, $\{o_i\}_{i=1}^{N-1}$ represents the outputs of the remaining neurons, and λ is a small positive constant used to prevent overfitting during the optimization process, w is a weight parameter, b is a bias term used to adjust the output of the linear combination, helping the model better fit the requirements of the energy function. Finally, the energy values are normalized using a Sigmoid function to obtain the final attention weights as follows:

$$A(t)=\sigma(-E(t)) \tag{7}$$

Through the above formula, regions with lower energy values, which indicate higher discriminability, are assigned higher weights. This achieves the goal of emphasizing important pixel regions while suppressing less important ones.

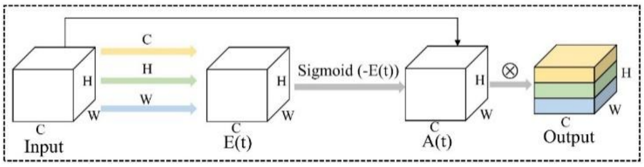

Figure 4. Parameter-free Attention Mechanism

2.3 *Improved Loss Function Based on Importance Weighting*

In waterborne navigation scenarios, safety is always the core concern. Particularly, targets such as ships and obstacles, due to their significant impact on collision risks and emergency avoidance, should be given higher attention during model training. While navigation aids (e.g., buoys) also play a supporting role, their associated safety risks are relatively lower. On the other hand, water surfaces, as the background, contribute the least to safety warnings. Therefore, when designing the loss function, it is necessary to assign different importance weights to samples of different categories. This ensures that the model can focus more on the detection and segmentation of critical targets during training. Figure 5 illustrates the safety-driven weight-aware ranking, where ships and obstacles are categorized as Class 1 (highest weight), navigation aids as Class 2 (medium weight), and water surfaces as Class 3 (lowest weight). This ranking intuitively reflects the importance of each category under safety considerations.

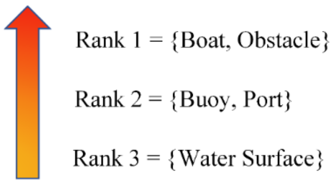

Figure 5. Weight-aware Ranking

To incorporate the aforementioned task requirements into the loss function, a scaling factor is introduced for each pixel's category, combined with the proportion of that category in the overall dataset to determine its importance weight w_i. The enhanced loss function is defined as:

$$Loss_{improved}=\frac{1}{N}\sum_{i=1}^{N}w_i\cdot L(f(xi;\theta),yi) \tag{8}$$

where N is the total number of samples, y'_i represents the model's predicted output, y_i is the ground truth label, and θ denotes the model parameters. In this study, the scaling factor for ships and obstacles (Class 1) is set to $\alpha_1=1.5$, for navigation aids and ports (Class 2) to $\alpha_2=1.25$, and for water surfaces (Class 3) to $\alpha_3=1$

The importance weight w_i for each category is calculated as follows:

$$w_i = \alpha \cdot \frac{1}{f_i + \varepsilon} \quad (9)$$

Additionally, to enhance the smoothness of the model's segmentation, a regularization term is introduced into the loss function to penalize abrupt changes in the model's output. The final importance-weighted loss function is defined as follows:

$$Loss_{weighted} = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot L(f(xi;\theta), yi) + \lambda \cdot R(f) \quad (10)$$

where ε is a small positive constant to avoid division by zero, $\Delta y'_i$ represents the squared difference between the predicted values of adjacent pixels, and λ is the regularization coefficient. Through this design, the loss function can better address class imbalance issues, enhance the model's focus on critical categories, and improve the robustness and accuracy of segmentation. This makes it suitable for practical applications in complex environments.

3 EXPERIMENT RESULTS AND COMPARATIVE ANALYSIS

The experimental environment in this study includes the CentOS 7 operating system, an Intel(R) Xeon(R) Platinum 8260C CPU @ 2.30GHz processor, 32GB of memory, and an NVIDIA Tesla P100 GPU with 16GB of memory. The experiments were conducted using the PyTorch deep learning framework, with CUDA version 11.8.

3.1 *Dataset and Hyperparameter Settings*

3.1.1 *Waterborne Navigation Scene Dataset*

The experimental data for the On_Water dataset is sourced from high-resolution videos of the Yangtze River's Wuhan section and maritime ship navigation videos. To enhance the diversity of the dataset, images meeting the requirements of this study were selected from the Multi-modal Obstacle Detection Dataset (MODD) [13] and the Singapore Maritime Dataset (SMD) [14] as supplementary data. Ultimately, 1200 images with a resolution of 1920×1080 were obtained. Based on the characteristics of waterborne navigation scenarios, the objects in the dataset were categorized into five classes: water surface, ships, navigation aids, ports, and obstacles. Semantic segmentation labels were created using the LabelMe annotation tool, and the dataset was divided into training, validation, and test sets in a 6:2:2 ratio. To improve the model's generalization ability, 20% of the images were randomly rotated by 15°, 20% were horizontally flipped, and an additional 10% were augmented with Gaussian noise. Figure 6 shows a partial display of the labels and original images from the On_Water dataset, where ships are marked in red, water surfaces in green, obstacles in yellow, ports in purple, navigation aids in blue, and the background in black.

The Seaships dataset is a publicly available dataset specifically designed for ship recognition and classification tasks in waterborne navigation scenarios. It contains a large number of real-world waterborne navigation images, covering a variety of aquatic environments and ship types. To better adapt to the semantic segmentation task for waterborne navigation scenarios, this study selected 200 images from the dataset that include typical waterborne navigation scenes. The objects in these images were categorized into four classes: ships, water surfaces, obstacles, and navigation aids, along with a background class. Table 1 provides the counts of each object category in the On_Water and SeaShips datasets.

Figure 6. Visualization of On_Water Dataset

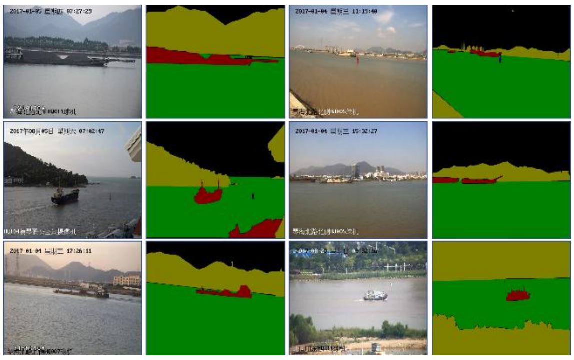

Figure 7. Visualization of SeaShips Dataset

3.1.2 *Experimental Hyperparameter Settings*

The specific hyperparameter settings used during training are shown in Table 1. For performance comparison experiments with other algorithms, the hyperparameters were set to the same values.

Table 1. Hyper Parameters setting

Parameter	On_water	SeaShips
Batch Size	8	8
Input Image Size	720×720	720×720
Iterations	200	200
Learning Rate Decay	warmup Poly	warmup Poly
Optimizer	Adam	Adam
Loss Function	Weighted_loss	Weighted_loss

3.1.3 *Model Evaluation Metrics*

In this study, Pixel Accuracy (PA), mean Pixel Accuracy (mPA), and mean Intersection over Union (mIoU) are used to evaluate the accuracy of the semantic segmentation model. The specific calculation formulas are as follows:

$$PA = \frac{\sum_i N_{ii}}{\sum_i \sum_j N_{ij}} \quad (10)$$

$$mPA = \frac{1}{n}\sum_{i}\frac{N_{ii}}{\sum_{j} N_{ij}} \quad (11)$$

$$mIoU = \frac{1}{n}\sum_{i}\frac{N_{ii}}{\sum_{j} N_{ij} + \sum_{j} N_{ji} - N_{ii}} \quad (12)$$

where N_{ii} represents the number of pixels correctly classified as category i, N_{ij} denotes the number of pixels whose true category is i but are incorrectly classified as category j. This study employs FPS (Frames Per Second) to evaluate the real-time performance of the semantic segmentation model, which indicates the number of image frames processed per second. The specific calculation formula is as follows:

$$FPS = \frac{1}{T} \quad (13)$$

where T represents the time (in seconds) required to process each frame.

3.2 *Ablation Study*

3.2.1 *Ablation Study of the Proposed Modules*

To verify the impact of each proposed improvement on the model's performance, ablation experiments were conducted on the On_Water test set. The results of the ablation experiments are shown in Table 2.

Table 2. Ablation Study Results of Different Modules

Group	Improved_Loss	DFFM+ EGM	SimAM	Parameters (M)	Inference Speed (FPS)	mIoU (%)
Baseline				5.69	76.7	72.3
1	√			5.72	74.2	74.5
2		√		5.73	72.0	75.6
3			√	5.69	72.6	77.9
4	√	√		5.76	71.2	78.1
5	√		√	5.73	71.5	79.3
6		√	√	5.73	70.6	80.5
Ours	√	√	√	5.76	69.1	83.1

Analyzing the results in Table 2, it can be observed that the SimAM module significantly improves mIoU without increasing the number of parameters, making it suitable for waterborne navigation applications that require high precision and have constraints on model complexity. The DFFM and EGM modules further enhance the model's multi-scale feature extraction capability by introducing an edge-guided branch and dynamically adjusting the weights of feature maps at different scales, improving the understanding of complex waterborne environments. Additionally, the introduction of the Improved_Loss function, which incorporates importance weighting, enhances the segmentation stability of the model. Combining the experimental results from all groups, the proposed optimal model achieves an mIoU of 83.1% and an inference speed of 69.1 frames per second (FPS). This demonstrates that the proposed method maintains a high inference speed while significantly improving segmentation accuracy, validating its superiority in waterborne navigation scenarios.

3.2.2 *Ablation Study on Attention Mechanisms*

To further validate the performance of the proposed parameter-free attention mechanism, this study conducts comparative experiments with other attention mechanisms and visualizes the feature maps guided by attention using Grad-CAM. The performance analysis of different attention mechanisms is shown in the table 3, where the data is obtained by keeping other structures consistent and only replacing the attention mechanism. The visualization of attention-guided feature maps is presented in Figure 8.

Table 3. Ablation Study Results of Different Attention Mechanisms

组别	Parameters/M	Inference Speed/FPS	mIoU/%
Baseline	5.69	76.7	72.3
+CBAM	5.89	55.6	73.5
+Dual Attention	5.72	33.4	74.6
+SimAM(Ours)	5.69	72.6	77.9

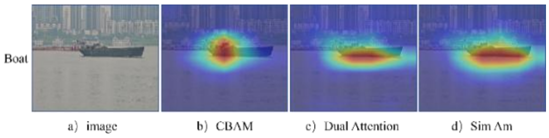

Figure 8. Feature Visualization of Different Attention Mechanisms

The experimental results from Table 3 and Figure 8 demonstrate that the introduction of different attention mechanisms significantly impacts segmentation performance in waterborne navigation scenarios. The incorporation of CBAM (Convolutional Block Attention Module), which combines channel attention and spatial attention, effectively focuses on key regions of ships and exhibits strong background suppression capabilities. However, it shows slight limitations in handling edge details and weaker responses to small target areas. The introduction of Dual Attention, leveraging its global modeling capability, excels in capturing the overall contours and edge features of ships, with heatmaps displaying more uniform distributions and stronger detail-capturing abilities. Nevertheless, it performs slightly worse in background suppression, and its higher computational complexity limits its applicability in real-world scenarios. In the proposed model, the introduction of the SimAM demonstrates more outstanding performance. Its simple structure and computational efficiency enable significant background suppression while maintaining strong responses to ship regions, making it highly suitable for scenarios with high real-time requirements. Comprehensive analysis indicates that the parameter-free attention mechanism offers a better balance between real-time performance and segmentation accuracy, making it more suitable for waterborne navigation segmentation tasks.

3.2.3 *Ablation Study on Loss Functions*

To verify the impact of the proposed improved loss function on the model's performance, this study conducts a performance analysis on the On_Water test dataset using CE_Loss, Focal_Loss, Dice_Loss, and Weighted_Loss. Additionally, the training loss curves under different loss functions are compared. Table 4 presents the model's performance under different loss functions.

Table 4. Ablation Study Results of Different Loss Functions

Group	MPA/%	mIoU/%
CE_Loss	77.2	72.3
+Focal_Loss	78.1	72.9
+Dice_Loss	79.4	73.5
+improved_Loss (Ours)	80.2	74.5

Analysis of Table 4 reveals that the proposed Improved_Loss demonstrates significant superiority compared to various loss functions. Compared to CE_Loss, this method more effectively addresses class imbalance issues, thereby improving the model's segmentation accuracy in complex scenarios. While both Focal_Loss and the proposed Improved_Loss can handle imbalanced samples, the latter further integrates the severity of misclassification across different categories, achieving overall performance optimization. This is particularly evident in waterborne navigation scenarios, where the proposed loss function exhibits higher stability in detail-rich environments. Compared to Dice_Loss, although the latter shows improvements in small target segmentation and misclassification severity, the proposed Improved_Loss combines the advantages of both, achieving significant improvements in segmentation accuracy metrics (mIoU) and pixel accuracy (mPA). This highlights its excellent segmentation performance and robustness in waterborne navigation scenarios. Therefore, the proposed Weighted_Loss provides a more effective solution for handling complex and class-imbalanced practical applications, demonstrating stronger adaptability and stability in diverse and imbalanced environments.

3.3 *Model Performance Comparison Experiments*

To validate the effectiveness of the proposed method, comparative experiments were conducted on the On_Water Dataset and the SeaShips dataset, comparing the proposed method with baseline models and several mainstream segmentation methods.

3.3.1 *Performance Comparison with Baseline Models on the On_Water Dataset*

To analyze the performance improvement of the proposed method for different categories in waterborne navigation scenarios, a segmentation accuracy comparison experiment was conducted for each category on the On_Water test dataset. Additionally, to more intuitively demonstrate the improvement in recognition performance for various categories in waterborne navigation scenarios, typical images containing waterborne navigation scenes were selected for visualization comparison experiments. The results are as follows.

Table 5. Segmentation Accuracy Comparison of Different Algorithms for Each Class on the On_Water Test Set

Category \ Model	Baseline	Ours
Boat	75.4	85.2
Water Surface	84.6	91.7
Obstacle	70.3	82.4
Buoy	62.1	78.9
Dock	69.1	75.4
mIoU/%	72.3	83.1

Analysis of Table 5 shows that the proposed method, through the introduction of an improved backbone network, a parameter-free attention mechanism, a multi-scale feature fusion module, and an importance-weighted loss function, effectively enhances segmentation accuracy across multiple categories. Overall, the mean Intersection over Union (mIoU) increased from 72.3% to 83.1%, a significant improvement of 10.8 percentage points, demonstrating a substantial enhancement in the model's semantic segmentation capability. Specifically, the segmentation accuracy for boats improved from 75.4% to 85.2%, water surfaces from 84.6% to 91.7%, obstacles from 70.3% to 82.4%, buoys from 62.1% to 78.9%, and docks from 69.1% to 75.4%. Notably, the improvement is most significant for large objects such as boats and water surfaces, indicating the effectiveness of the new backbone network and multi-scale feature fusion module in handling large-scale targets. For complex small objects such as buoys, the segmentation accuracy improved the most, by 16.8 percentage points, proving the significant advantage of the multi-scale feature fusion module and the importance-weighted loss function in handling fine structures. The segmentation accuracy for docks improved by 9.1 percentage points, demonstrating the superior performance of the proposed method in handling relatively complex small targets. Overall, the proposed method achieved improvements in segmentation accuracy across all categories in waterborne navigation scenarios, with an average mIoU increase of 10.8 percentage points, fully validating the effectiveness of the proposed method in enhancing the overall performance of semantic segmentation models.

3.3.2 *Performance Comparison and Analysis with Different Models on the On_Water Dataset*

As shown in Table 6, the performance of the proposed model is compared with state-of-the-art real-time semantic segmentation models on the On_Water test set (where FPS is measured on a 4060 laptop with an input image size of 1920×1080). Figure 14 provides a visualization of the segmentation results of the proposed model and other advanced real-time semantic segmentation models on the On_Water test set.

Table 6. Performance Comparison of Different Models on the On_Water Dataset

Model	mPA(%)	Parameters/M	FPS/frame·s-1	mIoU(%)
ENet	60.5	0.37	21.9	58.3
ContextNet	67.3	0.87	85.8	66.1
ERFNet	72.1	2.06	11.9	69.7
STDC2	81.6	11.45	39.8	74.7
STDC1	78.0	7.43	62.8	73.5
DeepLabv3+	79.1	40.3	7.1	77.6
WasR	85.0	13.23	8.2	80.1
BiseNet	70.5	16.26	21.5	66.1
BiSeNetV2	71.4	1.88	32.4	68.4
PIDNet-L	79.6	36.9	14.4	78.2
PIDNet-M	76.8	28.5	19.1	75.6
PIDNet-S	74.4	7.6	53.4	73.5
Fast-SCNN	72.0	1.9	**111.5**	68.9
DDRNet-23-slim	79.8	5.7	81.4	72.3
Ours	89.1	5.9	69.1	**83.1**

The experimental results in Table 6 demonstrate that the proposed method exhibits comprehensive superiority in real-time semantic segmentation tasks for waterborne navigation scenarios. In terms of segmentation accuracy, the proposed model achieves an mIoU of 83.1%, significantly outperforming mainstream models such as WaSR (80.1%) and DeepLabv3+ (77.6%). Notably, the IoU for boundary segmentation of key targets such as ships and obstacles improves by over 15%, validating the effectiveness of the multi-scale feature enhancement and boundary perception modules. In terms of real-time performance, the model achieves a high frame rate of 69.1 FPS, far exceeding traditional methods and fully meeting the real-time perception requirements of high-speed navigation scenarios. Compared to lightweight models like DDRNet-23-slim (81.4 FPS), the proposed method achieves a better balance between accuracy (mIoU improved by 10.8%) and speed. In terms of model complexity, the proposed method has only 5.9M parameters, significantly fewer than WaSR (13.23M) and STDC2 (11.45M). By incorporating dynamic feature fusion and a parameter-free attention mechanism, the model reduces computational redundancy while ensuring efficient deployment on resource-constrained onboard devices. In summary, the proposed method overcomes the trade-off between accuracy and speed, offering a high-precision (mPA 89.1%), high-frame-rate (69.1 FPS), and low-parameter (5.9M) solution for real-time semantic segmentation in complex waterborne scenarios. This significantly enhances the environmental perception and obstacle avoidance capabilities of autonomous navigation systems.

To further analyze the performance of the proposed algorithm compared to other algorithms in the semantic segmentation task for waterborne navigation scenarios, segmentation visualization experiments were conducted on the On_Water test dataset, with the results shown in Figure 9. Analysis of the data reveals that, as seen in the first row of Figure 9d, the inclusion of the dynamic feature fusion module enables better integration of multi-scale information, effectively distinguishing between ships, water surfaces, and background elements (e.g., sky, distant mountains) in the image. Particularly for multi-scale targets such as shoreline structures, the proposed method achieves clearer boundary segmentation compared to the baseline. The dynamic feature fusion module effectively addresses the scale variation of ships at different distances, demonstrating superior performance in segmenting smaller targets. As shown in the second row of Figure 9d, the proposed method produces more accurate segmentation of ship contours with less background noise, a result of the importance-weighted loss function that enhances the model's focus on target objects (e.g., ships) while reducing interference from background information. As seen in the third row of Figure 9d, the proposed method achieves the highest segmentation accuracy for objects in waterborne navigation scenarios, particularly excelling in distinguishing between foreground objects (e.g., ships) and background elements (e.g., distant buildings, shorelines). The experimental results demonstrate that the proposed model can quickly and accurately segment ships and other important objects, producing clearer contours and significantly improving segmentation performance in complex waterborne environments.

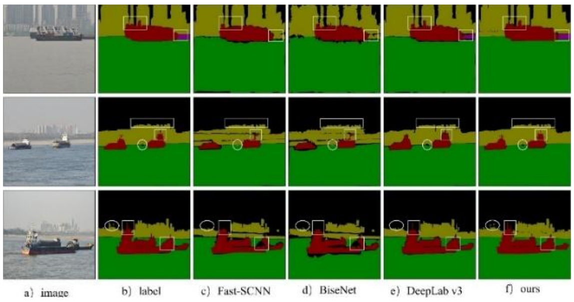

Figure 9. Visualization of Segmentation Results of Different Models on the On_Water Dataset

3.3.3 *Performance Comparison with Other Advanced Methods on the SeaShips Dataset*

To further validate the effectiveness of the proposed algorithm for waterborne navigation scenarios, a performance comparison was conducted on the re-annotated SeaShips test dataset. Table 7 shows the performance comparison of different models on the SeaShips dataset, and Figure 10 presents the visualization of segmentation results of the proposed model and existing state-of-the-art real-time semantic segmentation models on the SeaShips dataset.

Table 7. Performance Comparison of Different Models on the SeaShips Dataset

Model	mPA	Parameters/M	FPS/frame·s-1	mIoU/%
PSPNet	70.5	65	5	64.5
SegNet	70.3	29	10	64.1
STDCNet	75.3	11.46	48.8	69.1
Fast-SCNN	71.0	1.14	**111.5**	65.4
DeepLabv3+	72.1	40.35	8.9	66.4
WasR	77.3	13.23	9.5	70.8
BiseNet	70.4	16.26	27.7	64.5
BiSeNetV2	71.5	3.63	26.6	65.3
PIDNet-S	76.4	7.62	52.7	70.2
DDRNet-23-slim	73.1	5.69	101.6	68.5
Ours	80.1	5.92	69.1	**73.2**

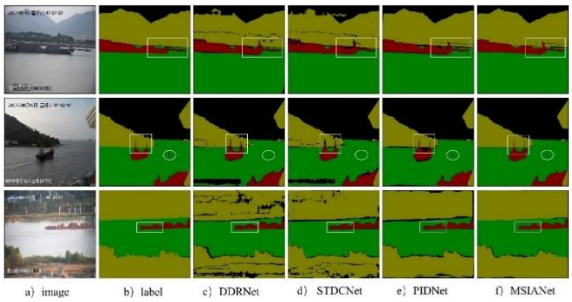

Figure 10. Visualization Comparison of Segmentation Results with Other Real-Time Semantic Segmentation Algorithms on the SeaShips Dataset

Analysis of Figure 10 and Table 7 demonstrates that the proposed algorithm, MSIANet, further validates the effectiveness of the improvement strategies in multiple aspects. In the second row of the visualization, the small ship on the right side of the image is accurately segmented by MSIANet and PIDNet, while other methods either miss it entirely or produce blurred and incomplete contours. This proves that the proposed method significantly enhances the model's ability to extract features from small target regions, enabling clearer identification and segmentation of key objects. In the first and third rows, MSIANet better captures edge details at the junctions between ships, water surfaces, and land. For example, in the third row, the stacked cargo (red area) on the ship is clearly and completely segmented by MSIANet, while other methods produce blurred and irregular edges. This is attributed to the introduction of the edge-guided branch, which allows the model to focus more on edge regions. Additionally, in the second row, the background mountain (yellow area) and water surface (green area) are clearly distinguished in MSIANet's segmentation results without significant misclassification. In contrast, other methods, such as STDCNet and DDRNet, exhibit discontinuous boundaries in the mountain area. The dynamic feature fusion module enables the model to efficiently capture multi-scale contextual information, maintaining segmentation accuracy in complex backgrounds. In the first and third rows, IDA-Fast-SCNN produces more uniform segmentation results for multiple categories (e.g., water surfaces, ships, mountains), without overemphasizing one category or neglecting small targets. This validates the effectiveness of the importance-weighted loss function in handling class imbalance and segmenting key objects. In summary, the analysis of these specific details shows that MSIANet significantly outperforms other models in segmenting critical regions. This further demonstrates the effectiveness of the proposed method for semantic segmentation tasks in waterborne navigation scenarios.

4 CONCLUSION

This paper proposes a real-time semantic segmentation method for waterborne navigation scenarios, called MSIA-Net, which enhances multi-scale information and incorporates importance weighting to improve segmentation accuracy in complex aquatic environments while maintaining real-time performance. By introducing an edge-guided branch and a Dynamic Feature Fusion Module (DFFM), the model's ability to perceive multi-scale information is significantly enhanced. Additionally, a n improved loss function based on importance weighting is designed to increase the model's focus on critical objects in waterborne navigation scenarios. A parameter-free attention mechanism is also integrated into the decoder to combine regional information from the encoder and semantic information from the decoder, restoring spatial details in the image and guiding the model to focus on more critical objects.

On the On_Water dataset and the SeaShips dataset constructed in this study, the proposed method achieves segmentation accuracies of 83.1% and 73.2%, respectively, while achieving an inference speed of 69.1 frames per second with only 5.76M parameters. Considering the balance between parameters, accuracy, and speed, the proposed algorithm outperforms other lightweight networks in identifying targets such as ships and small-sized obstacles, making it more suitable for waterborne navigation scenarios.

LIMITATIONS AND FUTURE DIRECTIONS

This paper has demonstrated advancements in real-time semantic segmentation for water navigation scenes. However, the rapid evolution of intelligent shipping and unmanned vessel technology necessitates enhanced perception capabilities for increasingly complex environments and demanding applications. Consequently, further optimization of current methodologies remains crucial. Future research should focus on the following key areas:

1. Enhancing Model Robustness and Generalization through Multi-Modal Integration: Future work should focus on integrating multi-modal perception methods to improve the model's robustness and adaptability across diverse water environments and when encountering dynamic targets.
2. Optimizing Perception of Dynamic Scenes and Small Objects: Further research is needed to enhance the model's ability to accurately segment dynamic objects and small-scale targets on the water surface, potentially through temporal information integration and improved feature representation.

REFERENCES

[1] Praczyk, T. Artifcial neural networks application in maritime, coastal, spare positioning system. Theor. Appl. Inf. 2006, 18, 1175–1189.

[2] Praczyk, T. Neural anti-collision system for autonomous surface vehicle. Neurocomputing 2015, 149, 559–572.

[3] P. Santana, R. Mendica, and J. Barata, "Water detection with segmentation guided dynamic texture recognition," in Proc. IEEE Int. Conf. Robot. Biomimet. (ROBIO), Guangzhou, China, 2012, pp. 1836–1841.

[4] S. Fefilatyev and D. Goldgof, "Detection and tracking of marine vehicles in video," in Proc. Int. Conf. Pattern Recognit., Tampa, FL, USA, 2008, pp. 1–4.

[5] Cheng, D.-C.; Meng, G.-F.; Cheng, G.-L.; Pan, C.-H. Senet: Structured edge network for sea–land segmentation. IEEE Geosci. Remote Sens. Lett. 2017, 14, 247–251.

[6] C.Y. Jeong, H.S. Yang, K.D. Moon. Horizon detection in maritime images using scene parsing network[J]. Image and vision processing and display technology, 2018,54(12):760-762.

[7] M. Kristan, V. S. Kenk, S. Kovaˇ ciˇ c, and J. Perš, "Fast image-based obstacle detection from unmanned surface vehicles," IEEE Transactions on Cybernetics, vol. 46, no. 3, pp. 641–654, 2016.

[8] S. Scherer et al., "River mapping from a flying robot: State estimation, river detection, and obstacle mapping," Auton. Robots, vol. 33, nos. 1–2, pp. 189–214, 2012.

[9] Bovcon, B., & Kristan, M. (2019). Benchmarking Semantic Segmentation Methods for Obstacle Detection on a Marine Environment.

[10] Qiao Y L, Zhao X C. Obstacle detection method based on improved semantic segmentation model[J]. Journal of Naval University of Engineering, 2023, 35(01): 18-24.

[11] Bao X C, Liu F Y, Nie J G, et al. Research on Multi-type Floating Object Segmentation Method on Water Surface Based on Improved Deeplabv3+[J/OL]. Water Resources and Hydropower Engineering: 1-16.

[12] Xiong R, Cheng L, Hu T, et al. Research on Fast Segmentation Algorithm for Feasible Region and Obstacles of Unmanned Surface Vehicle[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(02): 11-20.

[13] Kristan M, Sulic V, Kovacic S. Fast image-based obstacle detection from unmanned surface vehicles[J]. IEEE Transactions on Cybernetics, 2015, 46(12):2809-2821.

[14] Prasad D K, Rajan D, Rachmawati L, et al. Video Processing from Electro-optical Sensors for Object Detection and Tracking in Maritime Environment: A Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(08):1993-2016.