Enhancing Marine Pollution Detection in Remote Sensing via Self-Supervised Boundary Awareness

Shuaiyu Chen[1]
sc1321@exeter.ac.uk
Chunbo Luo[2]

Peng Ren[3]

Zeyu Fu*[1]
Z.Fu@exeter.ac.uk

[1] Multimodal Intelligence Lab,
Department of Computer Science.
University of Exeter,
Exeter, UK

[2] Department of Computer Science.
University of Exeter,
Exeter, UK

[3] College of Oceanography and Space Informatics.
China University of Petroleum (East China),
Tsingtao, China

Abstract

Accurate Marine Pollution Detection is challenging due to the vague, irregular, and low-contrast nature of pollutant boundaries. Existing boundary-aware remote sensing segmentation methods often rely on explicit boundary annotations or hand-crafted attention modules, limiting their effectiveness in marine environments where annotations are scarce and structures are complex. In this work, we introduce a fully Self-Supervised Boundary-Awareness (SSBA) block that can be seamlessly integrated into existing segmentation architectures for MPD. Our SSBA block integrates a VSS-based global extractor, a boundary-focused local extractor with deformable and frequency features, and an attention-guided fusion module to adaptively combine semantics and edges for boundary-aware prediction. To further enhance spatial sensitivity, we develop a boundary-aware attention module trained via boundary reconstruction, enabling dynamic focus on critical boundary regions. Experimental results on two marine pollution datasets show that our method consistently provides state-of-the-art performance, particularly under weak boundary conditions.

1 Introduction

Marine pollution, particularly oil spill detection from remote sensing imagery, poses a serious threat to marine ecosystems and demands timely and accurate monitoring [7, 22, 24]. Oil spills not only disrupt marine biodiversity but also cause long-lasting socio-economic consequences, underscoring the need for large-scale and reliable monitoring tools [1, 12]. With the increasing availability of SAR and multispectral imagery, semantic segmentation has become a core technique for detecting polluted regions, evolving from early thresholding and texture-based methods [2, 10, 31, 32] to modern deep learning models that provide stronger spatial–spectral representations and robustness across diverse conditions [23, 36, 39]. Nevertheless, delineating oil spill boundaries remains challenging due to their irregular morphology, blurred edges, and low contrast with the surrounding sea surface [19].

Existing boundary-aware segmentation methods attempt to mitigate these challenges by either relying on manual boundary annotations [20, 29] or introducing handcrafted attention mechanisms [25, 35]. However, boundary annotations are rarely available in large-scale marine datasets, and predefined attention designs often fail to adapt to the diverse and complex structures present in real-world conditions. These limitations substantially reduce the effectiveness of existing approaches, particularly under weak boundary cues and limited supervision [14]. This gap highlights the need for an annotation-free and adaptive strategy to strengthen boundary representation learning in marine pollution monitoring.

To address this issue, we reformulate boundary modelling as a self-supervised learning (SSL) task. The motivation for adopting SSL is to eliminate dependence on scarce and costly boundary annotations, which are particularly difficult to obtain in marine environments [16, 34]. Instead of manual labels, we introduce a fully self-supervised boundary-awareness (SSBA) block that can be seamlessly integrated into existing segmentation architectures for marine pollution detection, as illustrated in Fig. 1. The SSBA block incorporates a frequency-guided pseudo-boundary generator to derive structural priors, and a self-supervised boundary-aware attention module trained via boundary reconstruction, replacing conventional channel attention [26]. This design enables adaptive focus on boundary-critical regions and improves spatial sensitivity.

Extensive experiments on the M4D [19] and MADOS [17] datasets demonstrate that the proposed framework consistently outperforms strong baselines, particularly in scenarios characterised by vague boundaries and sparse annotations.

Our main contributions are as follows:

1. We propose a self-supervised boundary-aware block for marine pollution segmentation that removes the reliance on ground-truth boundary annotations.

2. We introduce a frequency-domain pseudo-boundary generator that provides structure-aware supervisory signals, together with a self-supervised boundary-aware attention module trained via boundary reconstruction to enhance spatial sensitivity.

3. We validate the effectiveness and generalisation of our block across two challenging marine pollution datasets, achieving consistent gains in both boundary accuracy and overall segmentation performance.

2 Related Works

Semantic segmentation has been extensively applied to remote sensing tasks, including marine pollution monitoring. Encoder–decoder architectures such as U-Net [27], LinkNet [3], DeepLab [5, 6], MixTransformer [4], and MambaUNet [33] have demonstrated strong performance in large-scale satellite imagery. Nevertheless, precise boundary localisation remains a persistent challenge, particularly in scenarios characterised by low contrast and vague structural boundaries. To address this issue, boundary-aware frameworks have been

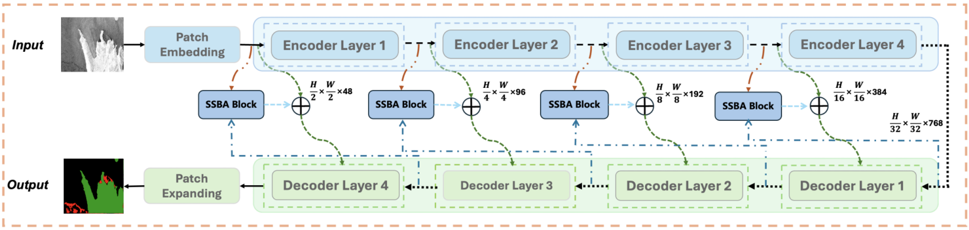

Figure 1: Illustration of integrating our Self-supervised Boundary-Awareness (SSBA) Blocks into the benchmark models for marine pollution detection.

developed to improve segmentation accuracy near object contours [11]. Common strategies include attention mechanisms [40] and boundary feature injection via skip connections [13]. These designs can enhance spatial feature alignment and gradient propagation but often rely on explicit boundary annotations or handcrafted priors. Consequently, they are prone to overfitting and may produce inconsistent boundary quality across scales [29]. Moreover, many existing approaches prioritise semantic coherence while overlooking fine-grained boundary discrimination [30]. Recent advances in self-supervised learning, such as reconstruction objectives, have shown promise in enhancing structural awareness without labels [15]. However, most of these methods treat boundary regions only implicitly, lacking dedicated mechanisms for boundary-level representation learning [9].

In contrast, our work explicitly targets boundary modelling without requiring annotations. We propose a frequency-guided pseudo-boundary generation mechanism, which derives structural priors from the frequency domain to guide boundary learning. In addition, we introduce a structure-aware attention module trained via boundary reconstruction loss, enabling the network to dynamically focus on informative boundary regions. Unlike handcrafted attention or auxiliary heads, our method achieves accurate boundary discrimination in a fully self-supervised manner.

3 Method

3.1 Problem Formulation

Marine pollution segmentation poses unique challenges due to the fragmented, low-contrast, and ambiguous nature of oil spill boundaries. To address these difficulties, our block explicitly models boundary-aware representations in a fully self-supervised manner. As shown in Fig. 1, the overall architecture follows an encoder–decoder paradigm, augmented with specialised branches that target different aspects of representation learning.

The Block is driven by two complementary inputs: (1) the original remote sensing image I, which provides semantic and appearance cues for segmentation; and (2) a frequency-derived pseudo-boundary prior E_{pseudo}, which introduces structural information about potential boundaries without requiring manual annotation.

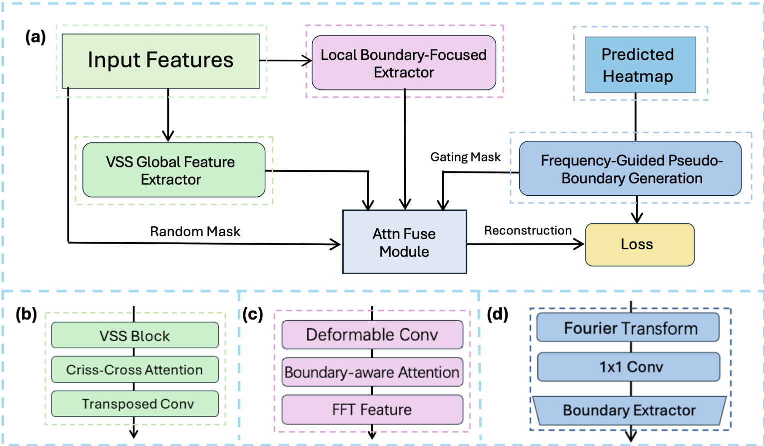

Figure 2: Overview of the proposed boundary-aware block. (a) Overall architecture. The input feature maps used in (b–d) are extracted from the corresponding encoder stage in Fig. 1 and fed into the subsequent modules. (b) VSS Global Feature Extractor for long-range semantics. (c) Local Boundary-Focused Extractor for fine-grained edges. (d) Frequency-Guided Pseudo-Boundary Generation providing structural priors. The predicted heatmap is produced from the output of the immediately preceding decoder layer.

3.2 Self-Supervised Boundary Awareness

3.2.1 Overview

As illustrated in Fig. 1, our block extends a standard encoder–decoder backbone with three complementary modules. The Local Boundary-Focused Extractor (LEFE) captures fine-grained spatial variations through multi-scale depthwise and dilated convolutions, thereby enhancing sensitivity to weak edge cues. The Global Feature Extractor (VSS) models long-range dependencies to preserve semantic coherence across regions. Meanwhile, the Frequency-Guided Pseudo-Boundary Generation (FGPBG) establishes the self-supervised learning task by deriving pseudo-boundaries from the frequency domain, which serve both as supervisory signals and as structural priors for feature integration. The outputs of these modules are adaptively fused by the Boundary-Guided Gating Fusion (BGGF), where pseudo-boundaries guide the balance between local detail and global context. In this way, the block achieves a unified design in which all modules contribute to a common objective of self-supervised boundary reconstruction(see Fig. 2 for detailed modules).

3.2.2 Local Boundary-Focused Extractor

Oil spill boundaries often appear weak or blurred, requiring the model to be highly sensitive to local intensity variations. As illustrated in Fig. 1(c), the Local Boundary-Focused Extractor (LEFE) is formulated as a multi-branch module that refines fine-grained boundary cues from complementary perspectives.

Given an input feature map $F \in \mathbb{R}^{H \times W \times C}$ from the encoder, three parallel operations are applied. A deformable convolution branch adaptively adjusts sampling positions to capture irregular boundary structures:

$$F_{\text{def}} = \phi\left(W_{\text{def}} *_{\text{def}} F\right), \tag{1}$$

where $*_{\text{def}}$ denotes deformable convolution and $\phi(\cdot)$ represents batch normalization followed by activation. In parallel, a boundary-aware attention operator enhances responses near semantic transitions:

$$F_{\text{att}} = \text{Attn}(F), \tag{2}$$

where $\text{Attn}(\cdot)$ denotes a learnable attention mask emphasizing edge-related pixels. In the frequency domain, a Fourier transform branch isolates high-frequency components correlated with object contours:

$$F_{\text{fft}} = \mathcal{F}^{-1}\big(H \cdot \mathcal{F}(F)\big), \tag{3}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are Fourier and inverse Fourier transforms, and H is a Gaussian high-pass filter.

The three outputs are concatenated and projected by a 1×1 convolution to form the local boundary-enhanced representation:

$$F_{\text{local}} = W^{(1\times 1)} \cdot \text{Concat}\big(F_{\text{def}}, F_{\text{att}}, F_{\text{fft}}\big). \tag{4}$$

The resulting $F_{\text{local}} \in \mathbb{R}^{H\times W\times C}$ retains detailed boundary information by integrating spatial adaptivity, attention-driven enhancement, and frequency-domain cues, while serving as a complementary input to the global pathway for subsequent fusion.

3.2.3 VSS Global Feature Extractor

While LEFE enhances local boundary detail, global semantic coherence is essential for maintaining region-level consistency. As illustrated in Fig. 1(b), the VSS Global Feature Extractor models long-range dependencies using the Visual State Space (VSS) model [21].

Given an encoder feature map $F \in \mathbb{R}^{H\times W\times C}$, we reshape it into a sequence $X \in \mathbb{R}^{T\times C}$ with $T = H \cdot W$ and apply the state-space update:

$$S_t = AS_{t-1} + BX_t, \qquad Y_t = CS_t + DX_t, \tag{5}$$

where A, B, C, D are learnable matrices. The outputs $\{Y_t\}_{t=1}^{T}$ are reshaped back to 2D and refined by a criss-cross attention block to strengthen global context. Finally, a transposed convolution restores spatial resolution, yielding the global feature representation:

$$F_{\text{global}} \in \mathbb{R}^{H\times W\times N}. \tag{6}$$

The resulting F_{global} complements the local representation F_{local}, ensuring semantic integrity across large regions while preserving scalability for high-resolution remote sensing imagery.

3.2.4 Frequency-Guided Pseudo-Boundary Generation

To overcome the lack of manual boundary annotations, we generate pseudo-boundary maps directly from the frequency domain. The key observation is that boundaries correspond to high-frequency components [38].

The Fourier transform of I is:

$$\hat{F}(u,v) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1} I(x,y)\, e^{-j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)}. \tag{7}$$

A Gaussian high-pass filter selects high-frequency components:

$$H(u,v) = 1 - \exp\left(-\frac{u^2+v^2}{2\sigma^2}\right), \tag{8}$$

giving the filtered spectrum:

$$\hat{F}_{\text{high}}(u,v) = H(u,v)\cdot\hat{F}(u,v). \tag{9}$$

The inverse FFT reconstructs a pseudo-boundary map:

$$E_{\text{pseudo}}(x,y) = \frac{1}{HW}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} \hat{F}_{\text{high}}(u,v)\, e^{j2\pi\left(\frac{ux}{H}+\frac{vy}{W}\right)}. \tag{10}$$

This E_{pseudo} serves dual purposes: (i) as a self-supervised reconstruction target for the boundary branch, and (ii) as a structural prior in the gating fusion, thereby linking frequency-derived point-level cues to regional feature integration.

3.2.5 Boundary-Guided Gating Fusion

To adaptively integrate local boundary detail and global semantic context, we design the Boundary-Guided Gating Fusion (BGGF) module. Given the local representation $F_{\text{local}} \in \mathbb{R}^{H\times W\times C}$, the global representation $F_{\text{global}} \in \mathbb{R}^{H\times W\times C}$, and the predicted boundary map $E_{\text{pred}} \in [0,1]^{H\times W\times 1}$, BGGF learns a spatial gating mask $M \in [0,1]^{H\times W\times 1}$ that controls the contribution of local versus global features at each pixel.

The fused feature representation is defined as:

$$F_{\text{fused}} = M \odot F_{\text{local}} + (1-M)\odot F_{\text{global}}, \tag{11}$$

where $\odot$ denotes element-wise multiplication. The gating mask is predicted by a lightweight convolutional sub-network:

$$M = \sigma\left(\text{Conv}_{1\times1}\left(\text{ReLU}\left(\text{Conv}_{3\times3}\left(\text{Concat}(F_{\text{local}}, F_{\text{global}}, E_{\text{pred}})\right)\right)\right)\right). \tag{12}$$

This mechanism enables the model to emphasise boundary-sensitive local cues in edge regions, while relying more on global semantics in homogeneous areas. As a result, the fused representation F_{fused} achieves a balanced encoding that preserves both fine boundary structures and large-scale contextual coherence.

3.3 Loss Functions

The overall training is guided by a dual-objective design: semantic segmentation and self-supervised boundary reconstruction:

$$L_{\text{total}} = L_{\text{seg}} + \lambda_1 L_{\text{rec}}, \tag{13}$$

Model	Per-class IoU (%)						Complexity	
	Sea	Oil	Look	Ship	Land	mIoU	Params (M)	Time (min)
U-Net [27]	93.90	53.79	39.55	44.93	92.68	64.97	30.94	4.47
U-Net*	92.12	**56.85**	34.32	49.53	89.98	**67.27**	32.05	5.16
LinkNet [3]	94.99	51.53	43.24	40.23	93.97	64.79	21.12	3.58
LinkNet*	94.10	**54.08**	40.02	44.18	92.00	**66.14**	22.21	4.11
PSPNet [37]	92.78	40.10	33.79	24.42	86.90	55.60	46.37	6.42
PSPNet*	93.41	**41.06**	31.58	25.23	87.92	**56.44**	47.45	7.21
DeepLabv2 [5]	94.09	25.57	40.30	11.41	74.99	49.27	47.48	6.37
DeepLabv2*	95.20	**27.15**	42.26	13.12	76.94	**50.44**	48.12	6.52
DeepLabv3+ [6]	96.43	53.38	55.40	27.63	92.44	65.06	42.88	5.17
DeepLabv3+*	96.10	**55.13**	56.61	29.48	93.55	**66.57**	43.97	6.48
MixTransformer [4]	93.33	40.23	34.06	24.55	87.48	56.52	27.06	7.08
MixTransformer*	94.32	**41.37**	35.69	25.95	88.41	**56.79**	28.16	8.34
MambaUNet [33]	93.48	40.78	34.31	25.32	87.56	56.17	26.51	5.03
MambaUNet*	94.29	**42.08**	34.80	25.84	88.68	**57.05**	27.63	5.59

Table 1: Per-class segmentation results and complexity comparison on the M4D [19] dataset. * indicates with SSBA block.

where L_{seg} is a standard segmentation loss (e.g., cross-entropy or Dice loss), and L_{rec} is an L2 reconstruction loss defined as:

$$L_{\text{rec}} = \|E_{\text{pred}} - E_{\text{pseudo}}\|_2^2. \tag{14}$$

Here, E_{pred} denotes the predicted boundary map, while E_{pseudo} is the frequency-derived pseudo-boundary map. This formulation explicitly enforces the network to reconstruct pseudo-boundaries in a self-supervised manner, ensuring structural fidelity without requiring manual annotations.

4 Experiments

4.1 Datasets

M4D Dataset [19]. The M4D dataset was developed for marine oil spill detection using SAR satellite imagery. It contains 1002 training and 110 test images, with pixel-level annotations for five semantic classes: *sea surface*, *oil spill*, *oil spill artefact*, *ship*, and *land*. The dataset exhibits strong boundary vagueness, especially for minority classes such as oil spills.
MADOS Dataset [17]. The MADOS dataset is a large-scale benchmark for marine pollution detection, including oil spills and marine debris. It consists of 174 Sentinel-2 satellite images collected from 2015 to 2022, with approximately 1.5 million annotated pixels spanning 15 semantic categories such as *oil spills*, *marine debris*, *ships*, and *turbid waters*. Each image tile (2.4 km $\times$ 2.4 km) is atmospherically corrected using ACOLITE and labelled by multiple experts. The dataset captures diverse oceanic and weather conditions across global coastal regions, making it a valuable resource for training and evaluating marine pollution detection models.

4.2 Implementation Details and Evaluation Metrics

Our model is built using the PyTorch framework. In the M4D dataset, the training template we used follows the design in HTSM [28]. We trained the model using an SGD optimiser with a momentum of 0.9 and a weight decay of 1e-4. Additionally, we set the initial learning rate to 0.001 and employed a "Poly" decay strategy. All experiments were implemented on an NVIDIA L40 48GB GPU. The batch size was set to 8, with a maximum of 100 epochs. Similarly, in the MADOS dataset, we used the official training framework provided by [18]. In the M4D dataset, we use the mean Intersection over Union (MIoU) score over the union to evaluate the model performance. In the MADOS dataset, both MIoU and average F1 (Ave.F1) scores are used to evaluate the model performance. These two evaluation metrics are based on the confusion matrix, which contains four components: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). For each class, IoU is defined as the ratio of the intersection and union of the predicted and ground truth values.

4.3 Overall Results

To evaluate the effectiveness of our proposed SSBA block, we integrate it into several widely used backbone networks, including U-Net [27], LinkNet [3], PSPNet [37], DeepLabv2 [5], DeepLabv2 (MSC) [5], and DeepLabv3+ [6], and evaluate them the M4D dataset. All encoder backbones are initialised with weights pretrained on ImageNet [8]. Table 3 reports the quantitative results for each segmentation method in the M4D Dataset. We observe that backbones trained with our Boundary-Aware Representation Learning Block achieve consistent improvements in both oil spill IoU and overall mIoU, demonstrating their effectiveness in accurately segmenting marine pollution regions, particularly under challenging boundary conditions.

We further evaluate our approach on the MADOS dataset by integrating it with U-Net [27], SegNeXt [13], and MarNext [17]. Table 2 presents the evaluation results on the MADOS dataset, where our block also consistently improves performance across different backbone architectures. Models trained with our boundary-aware strategy achieve notable gains in both mIoU and F1 score compared to their vanilla counterparts, further validating the generalisation ability of our method on diverse marine pollution scenarios, demonstrating the effectiveness of structural guidance under complex coastal backgrounds and multi-class segmentation settings.

To further assess the efficiency of our boundary-aware module, we report parameter counts and training time per epoch in Table 1. The additional cost is minor, with parameters increasing by about +1.1M (less than 5% overhead) across all models and runtime growing only 10–15% on average. This demonstrates a favorable trade-off between accuracy and efficiency. Moreover, the module is lightweight and easily integrated into diverse architectures, from conventional CNNs (U-Net, LinkNet) to advanced designs (Mix-Transformer, MambaUNet), making it well-suited for large-scale marine monitoring systems where both accuracy and scalability are critical.

4.4 Ablation Study

To evaluate the effectiveness of our overall architecture and its three core components, we conduct an ablation study based on the U-Net backbone using the M4D dataset. As shown in Table 3, incorporating the Local boundary-Focused Extractor (LEFE), Visual State Space Global Feature Extractor (MGFE), and the AttnFuse Module consistently improves the segmentation performance over the U-Net baseline. Specifically, we replace LEFE and MGFE with standard Laplacian-based boundary extraction and conventional CNN-based global context modules, respectively. We also compare our gated fusion with simple weighted

Model	F1	mIoU	OA
U-net [27]	63.8	51.0	82.9
U-net*	65.6	51.7	83.2
SegNext [13]	60.6	49.2	86.6
SegNext*	62.3	52.2	82.3
MariNeXt [10]	70.6	59.2	81.6
MariNeXt*	72.4	53.9	82.9
LinkNet [3]	66.3	56.8	77.5
LinkNet*	70.0	57.9	79.2

Table 2: Overall performance comparison on the MADOS [17] dataset. * indicates that our framework was used for training.

Model	Sea	Oil	Look	Ship	Land	mIoU
Baseline	93.90	53.79	39.55	44.93	92.68	64.97
Baseline + LEFE	95.15	55.67	40.64	46.44	93.88	66.17
Baseline + VGFE	95.89	55.79	40.90	46.72	93.94	66.29
Baseline + BGGF	95.85	55.26	40.77	46.63	94.20	66.87
Baseline + LEFE + BGGF	95.97	56.15	41.05	47.08	94.21	66.89
Baseline + LEFE + VGFE	96.14	56.13	41.18	46.86	93.96	66.85
Baseline + VGFE + BGGF	96.02	55.95	41.23	47.04	94.30	66.91
Baseline*	92.12	56.85	34.32	49.53	89.98	67.27

Table 3: Ablation study on the M4D [19] dataset. The baseline in the table is Unet [27], and * indicates that our proposed framework was used for training.

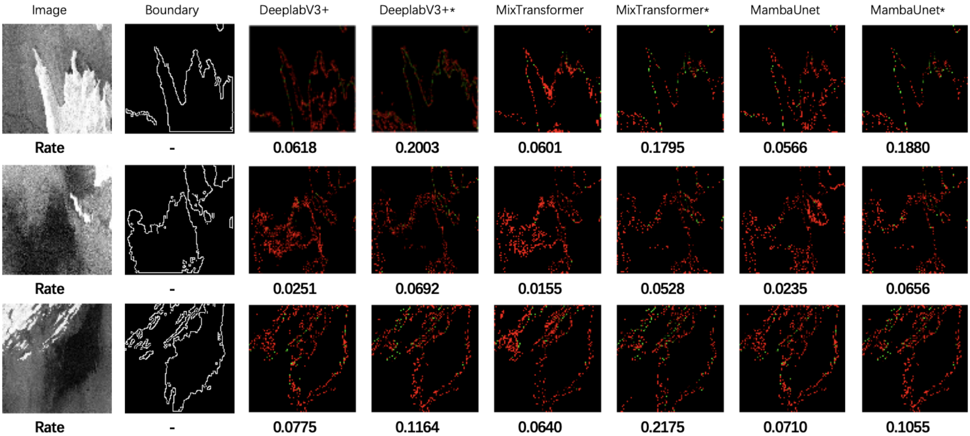

Figure 3: Boundary sensitivity comparison across different models. Each row visualises predictions of DeepLabv3+, MixTransformer, and MambaUNet, with quantitative scores indicating boundary precision. Models enhanced with our framework achieve consistently better boundary localisation.

addition to assess the contribution of each design. For clarity, we primarily focus on the IoU for oil spill regions and the overall mIoU. All three modules yield noticeable improvements over the baseline, demonstrating their effectiveness in enhancing boundary precision and overall segmentation accuracy.

4.5 Boundary Sensitivity Analysis

To further evaluate the model's capacity for fine-grained structural understanding, we conduct a boundary sensitivity analysis against baseline architectures. As shown in Figure 3, models enhanced with our block consistently generate boundaries that are sharper and closer to the ground truth compared with their vanilla counterparts. The boundary precision rate is defined as $TP/(TP+FP+FN)$. For example, DeepLabv3+ equipped with our block achieves a boundary precision of 0.2003, far surpassing the 0.0618 of the original model, while similar improvements are also observed for MixTransformer and MambaUNet. These results are particularly evident in challenging scenarios where oil spill edges appear blurry, fragmented, or indistinguishable from background sea clutter—conditions under which conventional models typically struggle. By contrast, our framework demonstrates the ability to suppress false boundaries while retaining true structural details, thereby reducing both missed detections and spurious responses. Importantly, the consistent improvements across

CNN- and transformer-based backbones highlight that the proposed boundary-aware representation learning is not tied to a specific architecture, but rather provides a general strategy for enhancing boundary precision in marine pollution segmentation tasks.

5 Conclusion

This paper presented a self-supervised boundary-aware framework for marine pollution segmentation, addressing the challenges of scarce boundary annotations and weak edge perception. The framework reformulates boundary modelling as a self-supervised reconstruction task, where the Frequency-Guided Pseudo-Boundary Generation module provides frequency-derived pseudo-labels and structural priors. Guided by this task, the Local Boundary-Focused Extractor (LEFE), Global Feature Extractor (VSS), and Boundary-Guided Gating Fusion (BGGF) jointly form an architecture aligned with boundary reconstruction. Experiments on multiple datasets demonstrate that our approach effectively enhances boundary delineation and overall segmentation accuracy under weak or ambiguous edges. The lightweight and modular design further enables seamless integration with diverse backbones, and future work will extend the framework to multimodal and semi-supervised settings.

References

[1] Werner Alpers, Benjamin Holt, and Ken Zeng. Remote sensing of oil spills: past, present, and future. *Remote Sensing of Environment*, 201:196–209, 2017.

[2] Camilla Brekke and Anne HS Solberg. Oil spill detection by satellite remote sensing. *Remote sensing of environment*, 95(1):1–13, 2005.

[3] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE, 2017.

[4] Jie-Neng Chen, Shuyang Sun, Ju He, Philip HS Torr, Alan Yuille, and Song Bai. Transmix: Attend to mix for vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12135–12144, 2022.

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[7] Saeid Dehghani-Dehcheshmeh, Mehdi Akhoondzadeh, and Saeid Homayouni. Oil spills detection from sar earth observations based on a hybrid cnn transformer networks. *Marine Pollution Bulletin*, 190:114834, 2023.

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, June 2009.

[9] Merv Fingas and Carl Brown. Review of oil spill remote sensing. *Marine pollution bulletin*, 83(1):9–23, 2014.

[10] B Fiscella, A Giancaspro, F Nirchio, P Pavese, and Paolo Trivero. Oil spill detection using marine sar images. *International Journal of Remote Sensing*, 21(18):3561–3566, 2000.

[11] Shin Fujieda, Kohei Takayama, and Toshiya Hachisuka. Wavelet convolutional neural networks. *arXiv preprint arXiv:1805.08620*, 2018.

[12] Maria Gonzalez et al. Oil spill impact on marine ecosystems: a review. *Marine Pollution Bulletin*, 110(1):6–18, 2016.

[13] Meng-Hao Guo, Cheng-Ze Lu, Qibin Hou, Zhengning Liu, Ming-Ming Cheng, and Shi-Min Hu. Segnext: Rethinking convolutional attention design for semantic segmentation. *Advances in Neural Information Processing Systems*, 35:1140–1156, 2022.

[14] Henning Heiselberg. Ship-iceberg classification in sar and multispectral satellite images with neural networks. *Remote Sensing*, 12(15):2353, 2020.

[15] Fenlong Jiang, Maoguo Gong, Hanhong Zheng, Tongfei Liu, Mingyang Zhang, and Jialu Liu. Self-supervised global–local contrastive learning for fine-grained change detection in vhr images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–13, 2023. doi: 10.1109/TGRS.2023.3238327.

[16] Xudong Kang, Bin Deng, Puhong Duan, Xiaohui Wei, and Shutao Li. Self-supervised spectral–spatial transformer network for hyperspectral oil spill mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–10, 2023.

[17] K. Kikaki, I. Kakogeorgiou, I. Hoteit, and et al. Detecting marine pollutants and sea surface features with deep learning in sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210:39–54, 2024.

[18] Katerina Kikaki, Ioannis Kakogeorgiou, Ibrahim Hoteit, and Konstantinos Karantzalos. Detecting marine pollutants and sea surface features with deep learning in sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210:39–54, 2024.

[19] Marios Krestenitis, Georgios Orfanidis, Konstantinos Ioannidis, Konstantinos Avgerinakis, Stefanos Vrochidis, and Ioannis Kompatsiaris. Oil spill identification from satellite images using deep neural networks. *Remote Sensing*, 11(15):1762, 2019.

[20] Shan Li, Weihong Deng, and JunPing Du. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2852–2861, 2017.

[21] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. *Advances in neural information processing systems*, 37:103031–103063, 2024.

[22] Jingwu Ma, Renfeng Ma, Qi Pan, Xianjun Liang, Jianqing Wang, and Xinxin Ni. A global review of progress in remote sensing and monitoring of marine pollution. *Water*, 15(19):3491, 2023.

[23] Lingfei Ma, Yu Liu, Xiang Zhang, Yun Ye, Guizhou Yin, and Brian A Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019.

[24] Matthew J McCarthy, Hannah V Herrero, Stephanie A Insalaco, Melissa T Hinten, and Assaf Anyamba. Satellite remote sensing for environmental sustainable development goals: A review of applications for terrestrial and marine protected areas. *Remote Sensing Applications: Society and Environment*, page 101450, 2025.

[25] Miguel Moctezuma, Flavio Parmiggiani, and Ludwin Lopez Lopez. Measuring marine oil spill extent by markov random fields. In *Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2014*, volume 9240, pages 57–62. SPIE, 2014.

[26] Zequn Qin, Pengyi Zhang, Fei Wu, and Xi Li. Fcanet: Frequency channel attention networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 783–792, 2021.

[27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[28] Abhishek Ramanathapura Satyanarayana and Maruf A Dhali. Oil spill segmentation using deep encoder-decoder models. *arXiv preprint arXiv:2305.01386*, 2023.

[29] Brian R Silliman, Philip M Dixon, Cameron Wobus, Qiang He, Pedro Daleo, Brent B Hughes, Matthew Rissing, Jonathan M Willis, and Mark W Hester. Thresholds in marsh resilience to the deepwater horizon oil spill. *Scientific Reports*, 6(1):32520, 2016.

[30] Jimmy Smith, Shalini De Mello, Jan Kautz, Scott Linderman, and Wonmin Byeon. Convolutional state space models for long-range spatiotemporal modeling. *Advances in Neural Information Processing Systems*, 36, 2024.

[31] Anne HS Solberg, Camilla Brekke, and Per Ove Husoy. Oil spill detection in radarsat and envisat sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3): 746–755, 2007.

[32] Konstantinos Topouzelis. Oil spill detection by sar images: Dark formation detection, feature extraction and classification algorithms. *Sensors*, 8(10):6642–6659, 2008.

[33] Ziyang Wang, Jian-Qing Zheng, Yichi Zhang, Ge Cui, and Lei Li. Mamba-unet: Unet-like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*, 2024.

[34] Jin Xu, Haixia Wang, Can Cui, Baigang Zhao, and Bo Li. Oil spill monitoring of shipborne radar image features using svm and local adaptive threshold. *Algorithms*, 13 (3):69, 2020.

[35] Tong Yang. Dynamic assessment of environmental damage based on the optimal clustering criterion–taking oil spill damage to marine ecological environment as an example. *Ecological Indicators*, 51:53–58, 2015.

[36] Liangpei Zhang et al. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):22–40, 2019.

[37] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[38] Qiqi Zhu, Yanan Zhang, Ziqi Li, Xiaorui Yan, Qingfeng Guan, Yanfei Zhong, Liangpei Zhang, and Deren Li. Oil spill contextual and boundary-supervised detection network based on marine sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–10, 2022. doi: 10.1109/TGRS.2021.3115492.

[39] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[40] Ziyi Zhu, Yafeng Yan, Ruilin Xu, Yun Zi, and Jinyin Wang. Attention-unet: A deep learning approach for fast and accurate segmentation in medical imaging. *Journal of Computer Science and Software Applications*, 2(4):24–31, 2022.