

Article

A Comparative Study on Ocean Front Detection in the Northwestern Pacific Using U-Net and Mask R-CNN

Caixia Shao ^{1,†}, Dianjun Zhang ^{2,3,†} and Xuefeng Zhang ^{1,*}¹ Yazhou Bay Innovation Institute, College of Marine Science and Technology, Hainan Tropical Ocean University, Sanya 572022, China² School of Marine Science and Technology, Tianjin University, Tianjin 300072, China³ Sanya Oceanographic Laboratory, Sanya 572025, China

* Correspondence: xuefeng.zhang@hntou.edu.cn

† These authors contributed equally to this work.

Abstract

Ocean fronts play a vital role in modulating climate variability, driving material transport, and maintaining the stability of marine ecosystems. Therefore, accurate identification of ocean fronts is of great significance for marine environmental monitoring and resource management. This study focuses on the Northwestern Pacific region and conducts a systematic comparison between two representative deep learning models—U-Net and Mask R-CNN—for automated ocean front detection. The objective is to evaluate the adaptability and strengths of different network architectures in handling multi-scale features, complex background conditions, and boundary delineation, thereby providing a theoretical basis for model selection and application-specific deployment. Experimental results show that U-Net achieves superior spatial consistency in large-scale frontal segmentation, with an IoU of 0.81 and a Dice coefficient of 0.76, while maintaining relatively high computational efficiency. In contrast, Mask R-CNN demonstrates stronger boundary modeling capabilities in detecting small-scale fronts and handling heterogeneous backgrounds, achieving an IoU of 0.78 and a Dice score of 0.73, though at the cost of increased computational demand. Overall, U-Net is more suitable for broad-scale automatic detection of ocean fronts, whereas Mask R-CNN exhibits greater potential in complex scene recognition. Integrating the structural advantages of both models holds promise for further enhancing the stability and accuracy of frontal detection, thereby offering robust technical support for ocean remote sensing analysis and environmental forecasting.

Keywords: ocean fronts; U-Net; Mask R-CNN; deep learning

Academic Editors: Cheng Sun and Bruno Buongiorno Nardelli

Received: 8 December 2025

Revised: 7 March 2026

Accepted: 23 March 2026

Published: 31 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Ocean fronts are boundary regions in the ocean where physical properties such as temperature, salinity, and density change sharply, often appearing as narrow bands with intensified temperature gradients. As significant mesoscale features in ocean dynamics, fronts exert profound impacts on material transport [1], climate regulation, and the distribution of primary productivity, while also playing a crucial role in the stability of marine ecosystems. In the context of accelerating climate change, the evolution and structure of ocean fronts have increasingly influenced ocean circulation patterns, fishery distributions, and ecosystem health. Therefore, accurate detection of ocean fronts is vital for understanding climate system evolution, enhancing marine forecasting, and supporting evidence-based resource management.

However, ocean front detection remains challenging due to the complex and dynamic nature of the marine environment. Fronts often exhibit multiscale coexistence, fuzzy boundaries, and strong background noise [2,3], which severely limit the performance of traditional detection methods. Existing approaches mainly include gradient-based thresholding, edge detection, and physics-based models. While such methods can extract frontal features under idealized conditions, they rely heavily on fixed parameters and prior knowledge, making them less effective in dynamically evolving or noisy environments. For example, the global front detection algorithm based on SST and chlorophyll proposed by Belkin and O'Reilly [2] performs well in open oceans but suffers from high false positives and inaccurate boundary localization in coastal regions.

In recent years, deep learning techniques have achieved remarkable progress in image recognition and semantic segmentation, providing new solutions for intelligent detection of complex marine structures. Compared with traditional algorithms, deep neural networks can learn hierarchical, nonlinear features from raw data via end-to-end training, significantly improving their ability to model unstructured patterns. U-Net and Mask R-CNN represent two prominent deep learning architectures in this field.

U-Net was originally proposed by Ronneberger et al. [4] for biomedical image segmentation [5,6]. Its symmetric encoder–decoder architecture, combined with skip connections, enables high-fidelity reconstruction of spatial structures by reusing high-resolution encoder features during decoding; this improves thin-structure and edge recovery but may also retain high-frequency fluctuations when the input is noisy or low-contrast. The model has been widely adopted in remote sensing applications such as coastline extraction, building delineation, and vegetation classification. Mask R-CNN, introduced by He et al. [7], extends Faster R-CNN by incorporating a mask prediction branch. It utilizes Region Proposal Networks (RPN) and Feature Pyramid Networks (FPN) to encode multi-scale spatial features, allowing simultaneous object detection and pixel-level segmentation. This design excels in challenging scenarios with occlusions, blurry boundaries, and dense targets.

Li et al. [8] applied U-Net to detect ocean fronts in the East China Sea and Kuroshio Extension, demonstrating the model's ability to restore frontal axis morphology and capture fine boundary features. Zhang et al. [9] investigated the frontogenesis of North Pacific subtropical sea surface temperature fronts and analyzed their dynamical connection with the overlying atmosphere, providing important insights into the physical mechanisms governing ocean front variability. More recently, deep learning-based ocean-front studies have further emphasized efficiency and rapid extraction capabilities, such as the edge-detection-model-based front mapping proposed by Felt et al. [10] and the lightweight SQNet designed for fast ocean-front identification by Niu et al. [11]. More recently, deep learning-based ocean-front studies have further emphasized efficiency and rapid extraction capabilities, such as the edge-detection-model-based front mapping proposed by Felt et al.

Furthermore, U-Net has been applied to urban road extraction, while Mask R-CNN has shown effectiveness in forest and water body classification for natural resource monitoring.

This study aims to conduct a comparative analysis of U-Net and Mask R-CNN for automated ocean front detection in the Northwestern Pacific. Using a curated dataset of satellite-derived sea surface temperature (SST) images, the performance of both models is evaluated under varying spatial scales, structural complexity, and background noise conditions. The comparison covers multiple dimensions, including segmentation accuracy, boundary consistency, and computational efficiency. Building on the comparative findings, we further discuss the feasibility of hybrid, coarse-to-refined strategies that combine the complementary strengths of U-Net and instance-aware segmentation frameworks to potentially improve stability and robustness in real-world marine scenarios. The findings

offer practical insights for the application of deep learning in ocean remote sensing and contribute to the development of intelligent front detection techniques.

Based on this context, this study focuses on the Northwestern Pacific region and conducts a comparative analysis of two deep learning models—U-Net and Mask R-CNN—in the task of automatic ocean front detection. A remote sensing dataset primarily composed of sea surface temperature (SST) imagery is constructed to evaluate the adaptability and performance of both models under varying conditions of spatial scale, structural complexity, and background interference. The assessment encompasses a quantitative comparison of detection accuracy, spatial consistency, and boundary precision, along with an evaluation of computational efficiency and model deployment cost. The findings aim to provide theoretical support and practical guidance for advancing intelligent ocean front recognition in remote sensing applications.

2. Study Area and Data

2.1. Study Area and Data Source

This study focuses on the Northwestern Pacific region, covering latitudes 0–50° N and longitudes 100–150° E, as shown in Figure 1. Figure 1 illustrates the geographic extent of the study domain based on SST reanalysis coverage, where coastlines, marginal seas, and major oceanic circulation systems are indicated to clarify the environmental context of the analysis. The study area encompasses several key marginal and open-sea domains, including the Bohai Sea, Yellow Sea, East China Sea, South China Sea, and parts of the Western Pacific Ocean. It spans temperate, subtropical, and tropical climatic zones, characterized by the confluence of multiple water masses and current systems such as the Asian monsoon circulation and the Kuroshio Current and its extension. The region features a complex topographic setting with continental shelves, island arcs, deep ocean basins, trenches, and coral reef ecosystems, making it highly heterogeneous in hydrographic and thermal structures. Due to its dynamic oceanographic processes and frequent frontal activities, this region plays a critical role in modulating regional climate variability, biogeochemical cycling, and the evolution of marine ecosystems. Accordingly, it serves as an ideal testbed for evaluating the robustness and adaptability of deep learning models in ocean front detection.



Figure 1. Spatial extent of the study area in the Northwestern Pacific derived from SST reanalysis coverage.

The SST data used in this study were obtained from the GLORYS12V1 global ocean reanalysis product, distributed by the Copernicus Marine Environment Monitoring Service (CMEMS, https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/services (accessed on 1 October 2025)). Developed by Mercator Ocean International, GLORYS12V1 is a fourth-generation global reanalysis system that assimilates satellite observations, in situ measurements, and numerical simulations. It is based on the NEMO ocean circulation model and the LIM3/SAM2 data assimilation framework. The product provides global, daily, three-dimensional ocean temperature fields from January 1993 to December 2020, with a horizontal resolution of $1/12^\circ$ and 50 vertical layers, offering sufficient spatial and temporal granularity to resolve oceanic frontal structures.

Because the GLORYS12V1 dataset is a reanalysis product that assimilates satellite and in situ observations through data assimilation, it provides spatially continuous SST fields without cloud-induced gaps. Therefore, explicit cloud masking or gap-filling procedures were not required in this study.

The dataset is provided in NetCDF format and referenced to the WGS 84 coordinate system. For the purposes of this study, daily mean SST data covering the $100\text{--}150^\circ$ E and $0\text{--}50^\circ$ N domain were extracted. The high spatial resolution and temporal continuity of this dataset make it well-suited for training and validating deep learning models in a variety of oceanographic conditions, thereby ensuring the generalizability and credibility of the experimental results.

2.2. Data Preprocessing

To ensure the physical consistency, numerical stability, and semantic accuracy of model inputs, a systematic preprocessing pipeline was applied to the original remote sensing data. This process comprised four key stages: sample selection, numerical normalization, data augmentation, and high-quality mask annotation.

During sample screening, we computed the two-dimensional SST gradient field and selected scenes with prominent frontal signatures, followed by visual inspection to remove distorted images or scenes without clear frontal structures. The resulting dataset contains 512 SST images, covering multiple seasons and subregions within the Northwestern Pacific to ensure representative spatiotemporal variability.

To improve convergence stability during training and minimize the impact of inter-sample scale variation on model learning, Min-Max normalization was applied independently to each image, linearly mapping pixel values to the $[0, 1]$ interval. This ensured consistency in the input distribution and enhanced the physical interpretability of the training data.

Subsequently, to enhance the model's generalization capability under complex environmental conditions and varying observational scenarios, a multi-strategy data augmentation procedure was applied to the training samples. The augmentation techniques included geometric transformations such as random rotation, horizontal flipping, and scaling, which improve the model's robustness to spatial variability. Figure 2 presents a representative example of the augmentation process.

In the mask generation phase, considering the inherent ambiguity of ocean front boundaries and the continuous variation in spatial scales, this study employed the Labelme [12–14] annotation tool—primarily using the polygon tool—to manually delineate ocean front contours (Figure 3). By connecting the vertices of each polygon, labeled regions were created and stored as JSON files to generate the initial binary masks. To ensure annotation reliability, all samples were independently labeled by two trained annotators with an oceanographic background, followed by expert reconciliation of disagreements.

The inter-annotator agreement, quantified using Cohen's kappa coefficient, reached 0.82, indicating high labeling consistency.

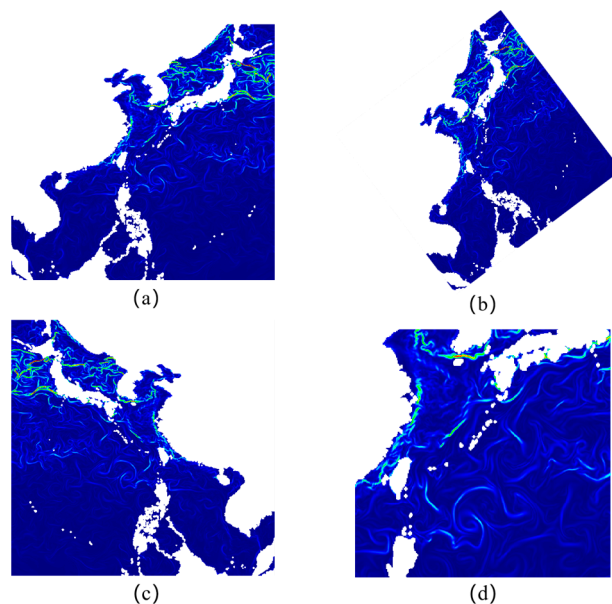


Figure 2. Ocean front sample (Green) and typical image augmentation strategies. (a) Original ocean front sample; (b) After random rotation; (c) After horizontal flipping; (d) After scaling.

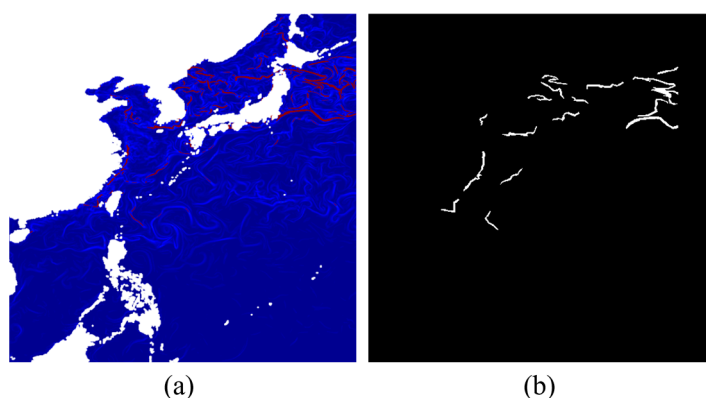


Figure 3. Illustration of ocean front annotation (Red) and label generation (White). (a) Ocean front sample in the original remote sensing image; (b) Corresponding binary mask generated via Labelme annotation.

This comprehensive preprocessing workflow substantially improved the quality and adaptability of the training data, providing a robust foundation for the effective training and reliable evaluation of deep learning models in ocean front detection.

3. Methodology

3.1. Model Architecture

The U-Net architecture, originally proposed by Ronneberger et al. [4], is a widely used fully convolutional network (FCN) designed for pixel-wise semantic segmentation tasks in medical and remote sensing imagery. As illustrated in Figure 4, the network adopts a symmetric encoder–decoder structure consisting of a contracting path and an expanding path. In this study, a vanilla U-Net is used, with a four-stage encoder–decoder and skip connections for multi-scale feature fusion. The input is a single-channel SST image, and the output is a single-channel binary front mask.

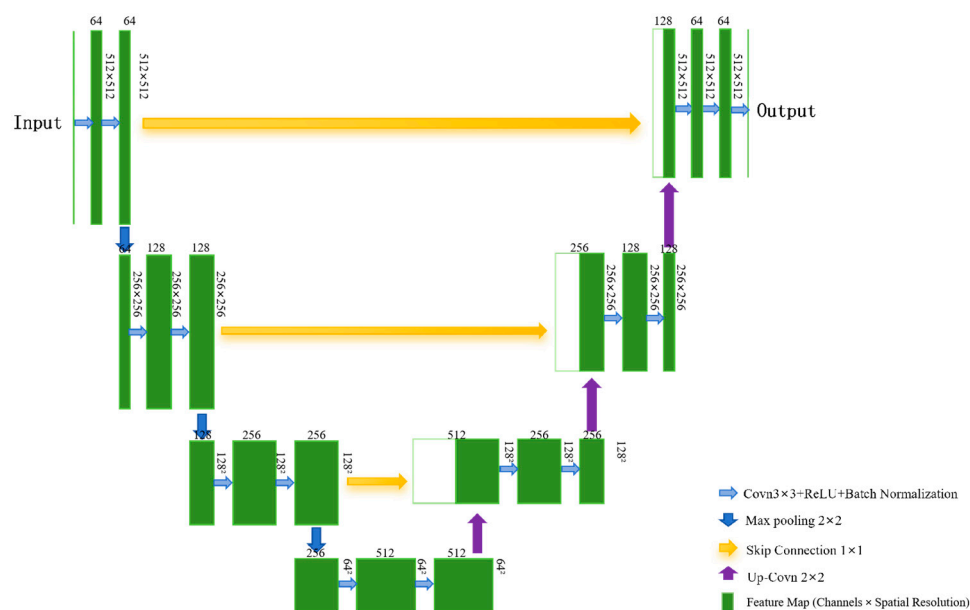


Figure 4. Schematic Diagram of the U-Net Architecture.

In the contracting path, each stage comprises two consecutive convolution operations followed by a max-pooling downsampling operation. Mathematically, the feature transformation at layer l can be expressed as:

$$x_{l+1} = \text{UpConv}(x_l) + \text{Concat}(x_l^{\text{cnc}}, x_l) \quad (1)$$

where x_l^{cnc} denotes the encoder-side feature map, and Concat represents channel-wise concatenation. This architecture is particularly suitable for detecting elongated, continuous structures with vague boundaries, such as oceanic fronts.

Figure 5 illustrates the architectural layout of the Mask R-CNN model adopted in this study. The overall design builds upon the two-stage object detection framework of Faster R-CNN and extends it with an instance segmentation branch to enable pixel-level mask prediction. Structurally, the network consists of five convolutional modules, three pooling operations, two fully connected (FC) layers, and a classifier head, forming a multi-branch deep architecture capable of multi-scale feature extraction and region-level semantic modeling.

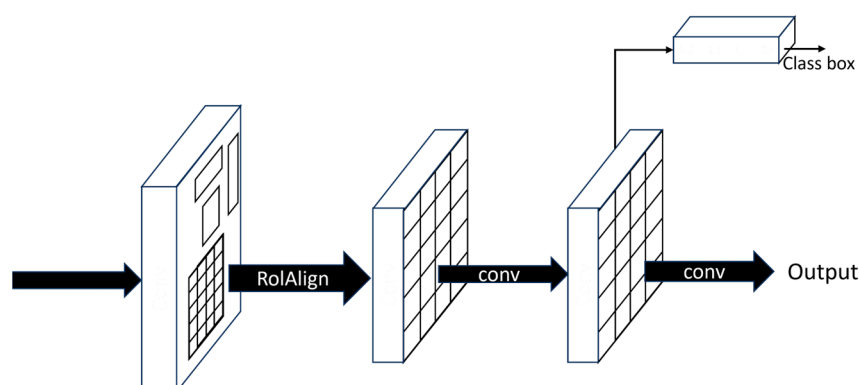


Figure 5. Architectural Layout of Mask R-CNN.

The input image is first processed by a backbone feature extractor—typically ResNet-50 or ResNet-101—combined with a Feature Pyramid Network (FPN), which together constitute the five hierarchical convolutional stages. These layers progressively encode semantic information at multiple spatial scales. Interleaved within the convolu-

tional stages, three max-pooling operations compress spatial dimensions while preserving key features, facilitating multi-scale representation alignment.

Following feature extraction, the Region Proposal Network (RPN) identifies candidate Regions of Interest (RoIs) and predicts their objectness scores and bounding box offsets. To improve spatial alignment, the model employs a RoIAlign operation, which replaces conventional RoIPooling and ensures sub-pixel precision in feature alignment, mitigating boundary mislocalization caused by quantization.

Subsequently, two parallel fully connected layers are introduced to perform classification and bounding box regression, respectively. In parallel, a lightweight fully convolutional mask branch is applied within each RoI to generate high-resolution binary masks, delineating object contours at the pixel level. The final classifier integrates outputs from all branches to produce the final instance-level predictions.

3.2. Model Training

To ensure a fair comparison of model performance, all training procedures were conducted under a unified configuration. The Adam optimizer was employed with an initial learning rate set to 1×10^{-4} , a batch size of 8, and a total of 100 training epochs. All SST inputs were resized to 512×512 and normalized to the range $[0, 1]$ using Min–Max scaling. The U-Net model adopted a single-channel input and generated a single-channel binary front mask. In addition, a weight decay of 1×10^{-4} was applied during optimization, and the learning rate was scheduled using StepLR with a decay factor gamma of 0.1 at a fixed step interval. An early stopping strategy was introduced to mitigate the risk of overfitting.

In this study, the design of the loss functions was tailored to the characteristics of each task. For U-Net, which targets full-image semantic segmentation, the loss function was defined as a weighted combination of Binary Cross-Entropy (BCE) [15,16] and Dice Loss, enabling a balance between global segmentation accuracy and boundary precision.

$$\mathcal{L}_{\text{U-Net}} = \alpha \cdot \text{BCE} + (1 - \alpha) \cdot \text{DiceLoss} \quad (2)$$

We set $\alpha = 0.5$ to equally weight BCE and Dice losses. The Dice Loss emphasizes the overlap in boundary regions, making it well-suited for imbalanced class distributions and fine-structure extraction. This helps enhance the model's ability to reconstruct fine-scale frontal features.

The training of Mask R-CNN involves three tasks: classification of candidate regions, bounding box regression, and pixel-level mask prediction. Each of these tasks corresponds to a distinct loss component, as previously described. This multi-task optimization strategy enhances the model's robustness in handling complex image regions and targets with indistinct boundaries. The overall loss function is formulated as a weighted sum of these task-specific losses.

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{box}} + \mathcal{L}_{\text{mask}} \quad (3)$$

The predefined training/validation/test split described in Section 2.2 was used for all experiments to ensure consistent evaluation.

3.3. Evaluation Metrics

To systematically evaluate the performance of U-Net and Mask R-CNN in ocean front detection over the northwestern Pacific, this study adopts a multi-dimensional evaluation framework covering segmentation accuracy, boundary consistency, detection sensitivity, and computational efficiency. These metrics provide a holistic assessment of model capabilities under varied spatial and dynamic oceanic environments.

3.3.1. Spatial Segmentation Accuracy

To quantify how well the predicted masks align with the ground truth, we use the Intersection over Union (IoU) [6,8,10,17,18] and the Dice Similarity Coefficient (Dice). These are defined as:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (4)$$

$$\text{Dice} = \frac{2|P \cap G|}{|P| + |G|} \quad (5)$$

where P denotes the predicted mask and G the ground truth. IoU evaluates the proportion of the intersected region over the union, focusing on overall area agreement, while Dice gives more weight to correctly predicted pixels in smaller regions, enhancing sensitivity to narrow frontal structures.

3.3.2. Boundary Adherence

To evaluate how accurately the models capture the fine-scale edges of fronts, we apply Hausdorff Distance (HD) and Contour Matching Score (CMS). The Hausdorff Distance is defined as:

$$HD(P, G) = \max \left\{ \sup_{p \in P} \inf_{g \in G} d(p, g), \sup_{g \in G} \inf_{p \in P} d(g, p) \right\} \quad (6)$$

where $d(p, g)$ represents the Euclidean distance between points on the predicted and ground truth boundaries. This metric reflects the worst-case boundary deviation. CMS, on the other hand, evaluates contour alignment by computing the average similarity between contours extracted from predicted and ground truth masks under multi-scale dilation kernels. Its exact form depends on the structural similarity formulation and is implemented via contour overlap matching. CMS is used to quantify contour-level alignment with multi-scale boundary tolerance. Specifically, we first extract one-pixel-wide contours from the predicted and ground-truth binary masks, denoted as $C(P)$ and $C(G)$. To account for slight boundary ambiguity, we apply morphological dilation to the contour maps with structuring elements of sizes $\{3 \times 3.5 \times 5.7 \times 7\}$. For each dilation scale k , a contour overlap score is computed as an IoU between the dilated contours. The final CMS is defined as the average IoU across all dilation scales.

3.3.3. Detection Sensitivity

Detection performance in complex backgrounds is evaluated using the F1-score, defined as the harmonic mean of precision and recall:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

where TP , FP , and FN denote true positives, false positives, and false negatives, respectively. F1-score is especially useful in identifying weak or fragmented frontal structures in high-noise SST scenes.

3.3.4. Computational Efficiency

To assess the models' feasibility for real-time or large-scale deployment, we report inference latency (average processing time per image) and GPU memory consumption during single-image prediction (batch size = 1). All experiments are conducted on an

NVIDIA RTX A5000 (24GB), ensuring consistency across comparisons. In addition, model size and computational complexity are reported to support deployment assessment. Model parameters are defined as the total number of trainable weights, and floating-point operations (FLOPs) are estimated based on the network architecture under the 512×512 single-channel SST input configuration used in this study. Under this setting, the U-Net model involves approximately 23 million parameters and about 36 GFLOPs per forward pass, whereas Mask R-CNN with a ResNet-50-FPN backbone involves approximately 48 million parameters and about 220 GFLOPs. These indicators complement latency and GPU memory statistics by providing a quantitative measure of computational demand and potential deployment cost during inference.

4. Results and Discussion

4.1. Recognition Results

To evaluate the performance of the models on a consistent testing dataset, the trained U-Net and Mask R-CNN architectures were applied to the task of ocean front detection, followed by both quantitative assessment and visual inspection. Figure 6 illustrates a side-by-side comparison of segmentation outputs from both models on the same test image, with subfigure (a) showing the prediction from U-Net and subfigure (b) depicting the result from Mask R-CNN.

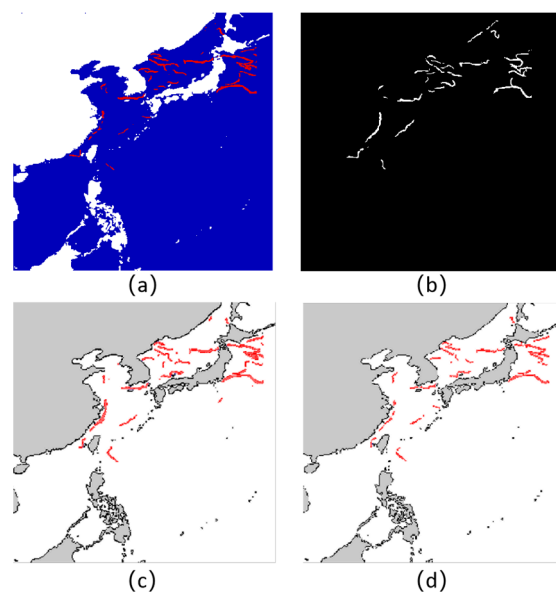


Figure 6. Qualitative comparison of ocean front detection results (Red). From left to right: (a) original SST field, (b) ground-truth annotation (White), (c) U-Net prediction, and (d) Mask R-CNN prediction.

The predictions were systematically evaluated using seven quantitative metrics. Table 1 summarizes the numerical comparison across all evaluation criteria, providing a comprehensive overview of the performance differences between U-Net and Mask R-CNN in the context of ocean front detection. To further illustrate model performance under representative oceanographic conditions, additional qualitative comparisons are presented in Figure 6. These examples include the original SST field, manually annotated ground truth, and the corresponding predictions from U-Net and Mask R-CNN. The visual comparison highlights differences in spatial continuity, boundary delineation, and sensitivity to weak frontal structures, providing intuitive support for the quantitative results summarized in Table 1.

Table 1. Quantitative comparison of segmentation performance between U-Net and Mask R-CNN.

Model	IoU (Mean [95% CI])	Dice (Mean [95% CI])	Hausdorff (Mean [95% CI])	CMS (Mean [95% CI])	F1-Score (Mean [95% CI])	Average Inference Time (ms)	GPU Memory Consumption (MB)
U-Net	0.81 [0.78, 0.84]	0.76 [0.73, 0.79]	4.2 [3.7, 4.7]	0.68 [0.65, 0.71]	0.79 [0.76, 0.82]	45.3	530
Mask R-CNN	0.78 [0.75, 0.81]	0.73 [0.70, 0.76]	2.7 [2.2, 3.2]	0.72 [0.69, 0.75]	0.82 [0.79, 0.85]	89.6	1100

To support uncertainty-aware marine monitoring, segmentation metrics are reported together with 95% confidence intervals derived from bootstrap resampling, and statistical significance between models was assessed using paired Wilcoxon signed-rank testing.

4.1.1. Segmentation Accuracy and Consistency

With regard to overall segmentation performance, U-Net achieved an Intersection over Union score of 0.81 and a Dice coefficient of 0.76. In contrast, Mask R-CNN scored 0.78 for Intersection over Union and 0.73 for Dice. The Intersection over Union metric primarily reflects the degree of area overlap between predicted and actual boundaries, whereas the Dice coefficient emphasizes agreement in fine-scale structures and is more sensitive to edge errors. The simultaneous improvement in both metrics suggests that U-Net can reconstruct broad frontal structures more precisely and maintain spatial continuity, particularly in open-ocean areas characterized by high frontal contrast. These results indicate a consistent advantage of U-Net in large-scale spatial continuity, while detailed statistical reliability metrics are summarized in Table 1. Paired Wilcoxon signed-rank testing confirms that the differences observed in both IoU and Dice metrics are statistically significant at the 0.05 level ($p < 0.05$), indicating that the performance gap between the two models is unlikely to be attributable to random variation.

4.1.2. Boundary Fitting and Geometric Consistency

Regarding boundary localization, Mask R-CNN demonstrates greater accuracy. Its Hausdorff distance, which measures the worst-case deviation between predicted and true contours, is significantly lower at 2.7 compared to 4.2 for U-Net. In addition, its Contour Matching Score reached 0.72, outperforming the 0.68 of U-Net. This suggests that Mask R-CNN achieves higher fidelity in reconstructing the geometric structure of frontal boundaries. These advantages can be attributed to the Region Proposal Network and Feature Pyramid Network, which together enable robust multi-scale boundary modeling in complex spatial contexts. The inclusion of a metric such as Mean Boundary Displacement would provide additional insight into the accuracy of continuous edge reconstruction.

4.1.3. Response to Weak Signals

The F1-score serves as a balance between precision and recall, offering an integrated view of detection quality. Mask R-CNN achieved an F1-score of 0.82, slightly higher than the 0.79 of U-Net. Its improved recall further indicates stronger responsiveness to low-contrast and fragmented frontal structures. This superior sensitivity is especially beneficial in coastal regions or scenes with significant background interference, where traditional methods often suffer from missed detections.

4.1.4. Computational Efficiency and Deployability

In terms of computational demand, U-Net exhibited substantially lower resource consumption. Its average inference time was 45.3 milliseconds, and GPU memory usage

was 530 megabytes. In contrast, Mask R-CNN required 89.6 milliseconds per inference and 1100 megabytes of memory. U-Net's lighter architecture and streamlined inference path make it more suitable for deployment in edge-based ocean monitoring systems. From a resource-performance perspective, U-Net delivers higher processing throughput per unit of memory, supporting broader scalability in large-scale remote sensing workflows.

These computational differences can be further interpreted in relation to the underlying architectural characteristics of the two models. In U-Net, skip connections help preserve fine-scale spatial information, which is particularly beneficial for detecting continuous but weak thermal gradients associated with ocean fronts. Meanwhile, the direct pixel-wise segmentation paradigm supports stable boundary continuity but may increase sensitivity to noise. In contrast, Mask R-CNN benefits from region proposal and multi-scale feature aggregation mechanisms that enhance boundary delineation in complex frontal environments.

This architectural complementarity suggests that the two models exhibit different strengths depending on application requirements. While Mask R-CNN generally provides clearer boundary delineation owing to its instance segmentation framework, its mask prediction branch operates at a relatively coarse spatial resolution compared with direct pixel-wise segmentation networks. Consequently, for thin or weak ocean fronts characterized by subtle thermal gradients, U-Net may exhibit superior pixel-level continuity despite slightly higher sensitivity to noise. Such differences highlight a practical trade-off between boundary sharpness, pixel-level consistency, and computational efficiency, indicating that model selection should consider both environmental complexity and operational deployment constraints.

4.1.5. Applicability and Scene Adaptability

Taken together, U-Net is better suited for large-scale frontal structures with clear boundaries, typically observed in mid-ocean settings. Conversely, Mask R-CNN shows greater robustness in complex, nearshore environments with overlapping and diffuse fronts. Further assessments involving time-series observations of front evolution, such as bifurcation, convergence, or migration, would help evaluate each model's adaptability in dynamic oceanographic conditions.

4.2. Discussion

This study addresses a practical question for marine monitoring: how U-Net and Mask R-CNN differ in detecting ocean fronts from SST imagery, and what trade-offs they imply for operational use. To complement these aggregate metrics with scenario-level interpretation, Figure 7 presents representative qualitative comparisons and highlights three boxed regions corresponding to distinct frontal conditions: Region A (blue) for open-ocean continuous fronts, Region B (yellow) for complex nearshore environments with heterogeneous backgrounds, and Region C (green) for weak-gradient or fragmented fronts.

Region A emphasizes open-ocean continuity, where the dominant requirement is spatially coherent reconstruction of extended frontal patterns. In this scenario, U-Net typically produces more continuous front structures, which is consistent with its fully convolutional encoder-decoder design that preserves global spatial organization and with skip connections that restore fine spatial details during upsampling. As a result, U-Net aligns well with monitoring tasks that prioritize stable delineation of large-scale fronts in open waters.

Region B highlights nearshore complexity, where background heterogeneity, coastline proximity, and overlapping structures place higher demands on boundary localization and shape adherence. In such conditions, Mask R-CNN tends to delineate boundaries

more clearly. This behavior is plausibly linked to its region-based proposal mechanism and multi-scale feature aggregation, where the Region Proposal Network and RoIAlign facilitate accurate localization and scale-invariant feature extraction. These characteristics are consistent with the boundary-focused metrics reported in Section 4.1 and provide an intuitive explanation for the improved contour adherence observed in complex coastal scenes. Nonetheless, the model's higher inference time and memory usage pose limitations for large-scale deployment in high-resolution satellite data processing.

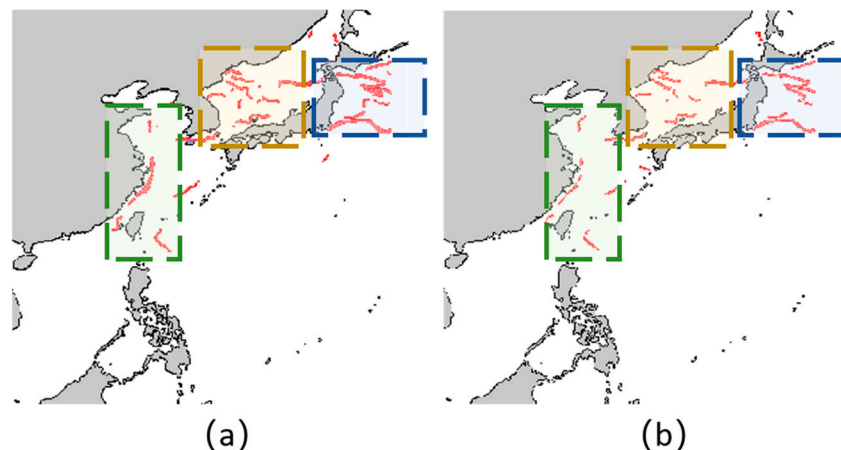


Figure 7. Ocean front detection results (Red) with annotated subregions (Region A (blue), B (yellow), C (green)). (a) U-Net prediction; (b) Mask R-CNN prediction.

Region C corresponds to weak-gradient or fragmented fronts, which represent a typical challenge in satellite-based front monitoring. Here, both models reveal characteristic trade-offs. U-Net may better maintain pixel-level continuity for thin and discontinuous fronts due to its dense pixel-wise formulation. However, direct fusion of high-resolution encoder features through skip connections can also amplify subtle texture fluctuations and sensor/background noise, occasionally leading to fragmented responses in low-contrast areas. Mask R-CNN, by contrast, can improve detection sensitivity to small or discontinuous frontal segments through its instance-aware pipeline, yet its mask branch operates at a comparatively coarser spatial resolution, which may lead to partial under-segmentation of very thin fronts.

Overall, the region-based comparisons in Figure 7 provide visual support for the main conclusions drawn from Section 4.1: U-Net favors spatial coherence and computational efficiency, whereas Mask R-CNN favors boundary localization and robustness in complex environments. This alignment with the study objective clarifies when each model may be preferable for marine monitoring workflows. Moreover, the complementary behavior motivates an application-oriented coarse-to-refined strategy, where U-Net is used to obtain coherent large-scale frontal structures and a subsequent refinement stage is applied selectively in nearshore or weak-signal regions to enhance boundary delineation when needed, balancing accuracy and deployment cost.

5. Conclusions and Future Perspectives

5.1. Conclusions

These findings indicate that the complementary characteristics of U-Net and Mask R-CNN may motivate future exploration of hybrid detection strategies, although such approaches are not implemented or validated in the present study. This study established a unified evaluation framework to systematically compare the performance of two representative deep learning models, U-Net and Mask R-CNN, in the task of ocean front detection in the Northwest Pacific. By incorporating multidimensional metrics and a high-quality

benchmark dataset, we comprehensively analyzed their capabilities in spatial segmentation, boundary alignment, fine-scale target recognition, and computational efficiency.

The results demonstrate that U-Net offers advantages in spatial consistency and computational economy, making it well-suited for large-scale, high-efficiency remote sensing workflows. In contrast, Mask R-CNN exhibited superior performance in boundary modeling and robustness in complex backgrounds, showing higher sensitivity to fine-scale frontal structures and better adaptability to multi-front, low-contrast environments.

Despite these promising findings, the study is subject to several limitations. The number of training samples remains constrained, and multi-source data fusion has not yet been explored. Future work may incorporate physical constraints, attention mechanisms, or semi-supervised strategies to further improve model generalization and physical interpretability. In particular, hybrid or multi-stage segmentation frameworks could be implemented by integrating physically meaningful variables (e.g., density gradients or frontal intensity indicators) derived from observational or reanalysis datasets as auxiliary input channels, introducing physics-informed regularization terms into the loss function to enhance dynamical consistency, and optionally applying a staged refinement strategy in regions with weak gradients or complex coastal structures. Such a structured integration of data-driven learning and physical constraints may help address boundary ambiguity and multi-scale frontal variability, while remaining within the scope of future methodological extensions rather than the present contribution.

5.2. Future Perspectives

Despite the promising outcomes demonstrated in this study, several challenges remain that warrant further exploration to advance deep learning-based ocean front detection.

First, oceanic fronts are characterized by diffuse boundaries and highly variable morphologies. Existing models still face difficulties in precisely localizing frontal boundaries, particularly in weak-gradient regions or areas with overlapping frontal structures. Future work should consider incorporating multi-scale feature fusion, boundary-aware modules, or structural priors to enhance the model's sensitivity to fine-scale edge details.

Second, the present study relies solely on sea surface temperature (SST) data. However, ocean fronts are often associated with concurrent variations in salinity, current velocity, and sea level anomaly (SLA). Integrating multi-modal remote sensing data, such as sea surface salinity (SSS) [14], geostrophic currents, and SLA, could improve model robustness in thermally indistinct regions and enhance physical consistency.

Lastly, to address the scarcity of labeled data, transfer learning, domain adaptation, and semi-supervised strategies offer promising avenues for extending model applicability across different oceanic regions and seasons. Such approaches can significantly improve generalization and adaptability under diverse environmental conditions.

In summary, future efforts should focus on optimizing model structures, integrating multi-source data, enabling real-time deployment, and enhancing cross-domain generalization. These directions will further solidify the role of deep learning in advancing oceanographic research and supporting sustainable marine resource management. Beyond methodological advances, improved front detection capability also has important oceanographic implications. Reliable delineation of sea surface temperature fronts facilitates the investigation of frontal variability associated with the Kuroshio system and monsoon-driven circulation, including seasonal migration, front intensification, and interaction with mesoscale eddies. These processes are closely linked to upper-ocean material transport, biological productivity, and regional biogeochemical cycling. Therefore, enhanced front mapping may provide valuable observational constraints for ocean dynamic studies and ecosystem-related analyses.

Author Contributions: Conceptualization, D.Z. and X.Z.; methodology, D.Z., X.Z. and C.S.; software, D.Z.; validation, C.S., D.Z. and X.Z.; formal analysis, D.Z.; investigation, D.Z.; resources, X.Z.; data curation, C.S.; writing—original draft preparation, D.Z.; writing—review and editing, C.S. and D.Z.; visualization, D.Z.; supervision, X.Z.; project administration, D.Z.; funding acquisition, X.Z., C.S. and D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (No. 2023YFC3107901), Hainan Province Science and Technology Special Fund (No. SOLZSKY2025007), Natural Science Foundation of Hainan Province (No. 425QN340), Scientific Research Foundation of Hainan Tropical Ocean University (No. RHDRC202335), Key Research and Development Program of China (No. 2023YFC3107701), and National Natural Science Foundation of China (No. 42375143).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The SST data used in this study were obtained from the GLORYS12V1 global ocean reanalysis product, distributed by the Copernicus Marine Environment Monitoring Service (CMEMS, https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/services (accessed on 1 October 2025)).

Acknowledgments: We thank anonymous reviewers for their comments and suggestions, which have improved the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Woodson, C.B.; Litvin, S.Y. Ocean fronts drive marine fishery production and biogeochemical cycling. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1710–1715. [CrossRef] [PubMed]
2. Belkin, I.M.; O'Reilly, J.E. An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. *J. Mar. Syst.* **2009**, *78*, 319–326. [CrossRef]
3. Ullman, D.S.; Cornillon, P.C.; Shan, Z. On the characteristics of subtropical fronts in the North Atlantic. *J. Geophys. Res. Ocean.* **2007**, *112*. [CrossRef]
4. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Lecture Notes in Computer Science; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
5. Heidler, K.; Mou, L.; Baumhoer, C.; Dietz, A.; Zhu, X.X. HED-UNet: Combined Segmentation and Edge Detection for Monitoring the Antarctic Coastline. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [CrossRef]
6. Zhao, N.; Huang, B.; Zhang, X.; Ge, L.; Chen, G. Intelligent identification of oceanic eddies in remote sensing data via Dual-Pyramid UNet. *Atmos. Ocean. Sci. Lett.* **2023**, *16*, 100335. [CrossRef]
7. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: Venice, Italy, 2017; pp. 2980–2988. [CrossRef]
8. Li, Y.; Liang, J.; Da, H.; Chang, L.; Li, H. A Deep Learning Method for Ocean Front Extraction in Remote Sensing Imagery. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
9. Zhang, L.; Xu, H.; Ma, J.; Shi, N.; Deng, J. North Pacific subtropical sea surface temperature frontogenesis and its connection with the atmosphere above. *Earth Syst. Dyn.* **2019**, *10*, 261–270. [CrossRef]
10. Felt, V.; Kacker, S.; Kusters, J.; Pendergrast, J.; Cahoy, K. Fast Ocean Front Detection Using Deep Learning Edge Detection Models. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–12. [CrossRef]
11. Niu, R.; Tan, Y.; Ye, F.; Gong, F.; Huang, H.; Zhu, Q.; Hao, Z. SQNet: Simple and fast model for ocean front identification. *Remote Sens.* **2023**, *15*, 2339. [CrossRef]
12. Chen, C.-T.A. Chemical and physical fronts in the Bohai, Yellow and East China seas. *J. Mar. Syst.* **2009**, *78*, 394–410. [CrossRef]
13. Good, S.; Fiedler, E.; Mao, C.; Martin, M.J.; Maycock, A.; Reid, R.; Robert-Jones, J.; Searle, T.; Waters, J.; While, J.; et al. The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses. *Remote Sens.* **2020**, *12*, 720. [CrossRef]
14. Hickox, R.; Belkin, I.; Cornillon, P.; Shan, Z. Climatology and seasonal variability of ocean fronts in the East China, Yellow and Bohai seas from satellite SST data. *Geophys. Res. Lett.* **2000**, *27*, 2945–2948. [CrossRef]

15. Ruby, A.U.; Theerthagiri, P.; Jacob, I.J.; Vamsidhar, Y. Binary cross entropy with deep learning technique for Image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 5393–5397. [[CrossRef](#)]
16. Jadon, S. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Via del Mar, Chile, 27–29 October 2020*; IEEE: New York, NY, USA, 2020; pp. 1–7. [[CrossRef](#)]
17. Solórzano, J.V.; Mas, J.F.; Gao, Y.; Gallardo-Cruz, J.A. Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery. *Remote Sens.* **2021**, *13*, 3600. [[CrossRef](#)]
18. Ulmas, P.; Liiv, I. Segmentation of Satellite Imagery using U-Net Models for Land Cover Classification (Version 1). *arXiv* **2020**, arXiv:2003.02899. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.