remote sensing

MDPI

Article

A Deep Learning Framework for Automatic Detection and Segmentation of Internal Solitary Waves in Satellite Imagery over the Northern South China Sea

Guocai Rong [1], Huanhe Dong [1], Yuan Kong [1], Ze Liu [1,2,3,4] and Yong Fang [1,*]

[1] College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China; rongguocai96@sdust.edu.cn (G.R.); donghuanhe@sdust.edu.cn (H.D.); kongyuan@sdust.edu.cn (Y.K.); liuze@qdio.ac.cn (Z.L.)
[2] Key Laboratory of Ocean Observation and Forecasting and Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China
[3] Laboratory for Ocean Dynamics and Climate, Qingdao Marine Science and Technology Center, Qingdao 266237, China
[4] Laoshan Laboratory, Qingdao 266237, China
* Correspondence: fangyong@sdust.edu.cn

Highlights

What are the main findings?

- YOLO11-ISW integrates DSConv and CBAM to improve the detection and segmentation of curved and low-contrast ISWs.
- Application to 2011-2020 MODIS imagery reveals pronounced spatial clustering and strong seasonal and lunar-cycle variability of satellite-observed ISWs in the northern South China Sea.

What are the implications of the main findings?

- The proposed framework provides an efficient approach for automated ISW monitoring, offering reliable information for marine environmental observation and oceanographic applications.
- Georeferenced crest-line extraction provides a foundation for future quantitative analysis of ISW propagation characteristics using suitable multi-temporal observations.

Abstract

Internal solitary waves (ISWs) play an important role in ocean energy transport and mixing, yet their elongated, curved crest lines and low contrast make automated extraction from satellite imagery challenging. To address these challenges, this study proposes YOLO11-ISW, an improved YOLO11-based framework for joint ISW detection and pixel-level segmentation. Dynamic Snake Convolution (DSConv) is introduced to enhance the representation of curved crest-line structures, while the Convolutional Block Attention Module (CBAM) strengthens responses to low-contrast ISW features. A MODIS-based ISW dataset from the northern South China Sea was constructed for model training and evaluation. YOLO11-ISW achieved detection Precision, Recall, and F1-score values of 92.01%, 80.73%, and 86.00%, respectively, and segmentation values of 93.14%, 82.60%, and 87.55%, respectively. The model achieved the highest detection and segmentation F1-scores among the evaluated methods. The segmentation results were further georeferenced to extract ISW crest-line locations, and the framework was applied to MODIS imagery from 2011 to 2020 to investigate satellite-observed ISW spatiotemporal variability. The detected ISWs exhibited pronounced spatial clustering, strong spring–summer seasonality, and clear lunar-cycle modulation, consistent with previous observations in the northern South China

Check for updates

Academic Editor: Chung-Ru Ho

Received: 7 October 2026
Revised: 5 November 2026
Accepted: 6 November 2026
Published: 8 November 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) license.

Sea. These results demonstrate the potential of YOLO11-ISW for automated, long-term, and large-scale monitoring of ISWs using satellite remote sensing imagery.

Keywords: internal solitary waves; MODIS; deep learning; detection and segmentation; spatiotemporal distribution

1. Introduction

ISWs are nonlinear wave phenomena generated within the ocean interior under the influence of stratification and external forcing. They are widely distributed throughout the global oceans [1]. During propagation, large-amplitude ISWs can induce strong shear and turbulence, posing significant risks to offshore structures, submarine navigation, and marine engineering activities [2,3]. Meanwhile, ISWs transport nutrient-rich deep water into the upper ocean, promoting phytoplankton growth and thereby playing an important role in material cycling and energy transfer within marine ecosystems [4,5]. Therefore, accurate and automated detection of ISWs and characterization of their spatiotemporal variability are of considerable scientific and practical importance for marine engineering safety and the study of ocean dynamical processes.

Early ISW observations relied primarily on in situ instruments, such as moorings and expendable bathythermographs. Although these observations provide valuable information on ISW dynamics, their limited spatial coverage and relatively high deployment costs restrict large-scale monitoring [6,7]. With the rapid development of satellite remote sensing technology, sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Synthetic Aperture Radar (SAR) have provided abundant observations of ISW-induced surface manifestations, greatly facilitating ISW detection and analysis [8].

With the increasing availability of satellite imagery, various remote sensing image processing approaches have been developed for ISW detection. Conventional automated approaches have mainly relied on wavelet transforms, spectral analysis, edge detection, and machine-learning classifiers. Ródenas and Garello [9] combined two-dimensional wavelet transforms with oceanographic knowledge for ISW localization, while subsequent studies incorporated support vector machines, FFT/CWT analysis, and edge-based feature extraction [10–12]. Although these methods improved the automation of ISW identification, their performance remains dependent on manually designed features and can be sensitive to complex backgrounds and image noise.

Recently, deep learning methods have demonstrated strong capabilities in automatic feature extraction and have been increasingly applied to ISW recognition from remote sensing imagery [13,14]. For instance, Faster R-CNN-based approaches have been developed for automatic detection of internal waves in SAR images [15]. Furthermore, improved YOLO-based models have been proposed by incorporating advanced convolution and attention mechanisms to enhance the extraction of irregular internal wave signatures [16,17]. Although object detection methods can efficiently localize ISWs, bounding-box representations cannot preserve the detailed crest-line morphology required for pixel-level geospatial characterization.

To obtain pixel-level representations of ISWs, further studies have focused on segmentation techniques [18]. Ma et al. [19] proposed a two-stage algorithm that first used decision fusion to identify regions containing internal waves in SAR images and then applied an improved Pixel Attention U-Net (PAU-Net) model for stripe segmentation. This approach enabled the automatic and accurate extraction of ocean internal wave features from full-scene SAR imagery of the Sulu Sea. Subsequently, an automatic segmentation method for

ocean internal wave stripes was developed based on the SegFormer framework [20]. The proposed method enabled efficient and accurate identification and localization of internal waves in SAR imagery, and experimental results demonstrated its superior performance compared with conventional segmentation models. In addition, the IWE-Net model, based on deep convolutional neural networks, was developed specifically for the automatic detection and extraction of internal wave features from both SAR and optical satellite imagery [13]. More recently, a few-shot learning method based on pyramid conditional generative adversarial networks was designed [21]. This method enabled the precise extraction of ocean internal wave stripes using only four labeled images, effectively alleviating the problem of limited training data for internal wave detection in SAR imagery.

Overall, deep learning approaches have substantially improved the efficiency and accuracy of automatic ISW recognition in remote sensing imagery. Object detection methods can rapidly locate ISWs in large-scale scenes, whereas semantic segmentation methods are capable of providing pixel-level fine structural information. Nevertheless, several limitations remain in existing studies. First, ISWs generally exhibit elongated and curved crest line morphologies, whereas conventional convolution operations are more effective at extracting regular geometric features and have limited capability in representing complex curved structures. Second, the contrast between ISWs and the surrounding ocean background is often low, which may lead to missed detections or inaccurate segmentation boundaries. Furthermore, most existing studies focus primarily on detection or segmentation performance, with limited integration of recognition results with geospatial information for long-term and large-scale characterization of ISW spatiotemporal variability. Their outputs are typically limited to target localization or contour extraction, lacking an integrated framework that further utilises recognition results for the analysis of propagation characteristics, such as propagation velocity estimation.

The northern South China Sea is one of the most active regions for ISW generation due to its complex bathymetry and strong tidal forcing, which produce favourable conditions for frequent ISW occurrence [22,23]. These energetic ISWs can induce strong vertical currents, enhanced shear, and turbulence in the upper and intermediate ocean, potentially affecting the safety and operation of offshore engineering structures, marine transportation, and subsea infrastructure. Therefore, timely and accurate monitoring of ISW occurrence and propagation is of practical importance for marine engineering risk assessment and operational planning.

To address the limitations of existing studies, an improved YOLO11-ISW model is proposed to achieve integrated ISW object detection and pixel-level segmentation, enabling the simultaneous acquisition of both target locations and contour information. DSConv is introduced to enhance the representation capability of complex curved crest-line structures, while the CBAM is incorporated to improve the extraction of low-contrast ISW features. The segmentation outputs are further integrated with GeoTIFF georeferencing information to derive the geographic locations and crest-line morphology of detected ISWs. Finally, the proposed framework is applied to a decade-long MODIS archive to characterize the spatiotemporal distribution of satellite-observed ISWs in the northern South China Sea. This integrated framework provides an automated approach for ISW detection, segmentation, geospatial characterization, and long-term remote-sensing monitoring.

The structure of this paper is organized as follows: Section 2 describes the acquisition and preprocessing of MODIS imagery over the northern South China Sea, the construction and annotation of the ISW dataset, and the architecture and design of the proposed YOLO11-ISW framework. Section 3 presents the experimental evaluation of the proposed method, including ablation and comparative analyses, georeferenced ISW crest-line extraction, and the spatiotemporal distribution analysis of satellite-observed ISWs from 2011 to

2020. Section 4 discusses the main findings, applicability, and limitations of the proposed framework, while Section 5 summarises the principal conclusions and future perspectives.

2. Materials and Methods

2.1. Data

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key sensor operated by NASA aboard the Terra (launched in 1999) and Aqua (launched in 2002) satellites, with an orbital altitude of approximately 700 kilometers [24]. Through multispectral observations, MODIS continuously provides essential data products for global environmental and climate studies. MODIS imagery can be conveniently accessed and visualized through NASA's Worldview interactive online data visualization platform (https://worldview.earthdata.nasa.gov/, accessed on 7 October 2026). Additionally, Worldview offers corrected reflectance true-color imagery, in which oceanic ISWs typically appear as alternating bright and dark stripes [25]. GeoTIFF is a raster image format that embeds georeferencing information within the standard TIFF structure [26]. GeoTIFF links image pixels to geographic coordinates, enabling the segmentation results to be transformed into georeferenced ISW features for subsequent spatial and spatiotemporal analyses.

To ensure data consistency, corrected reflectance true-color MODIS imagery with a spatial resolution of 250 m from the northern South China Sea was uniformly adopted as the data source, and all images were exported in GeoTIFF format. The study area extends from 112°E to 122°E and from 17°N to 24°N. The satellite imagery of the study area is shown in Figure 1a, where the region outlined by the red box indicates the area in which ISWs frequently occur. The enlarged view in Figure 1b further illustrates that ISWs appear as distinct alternating bright and dark stripes in the imagery.

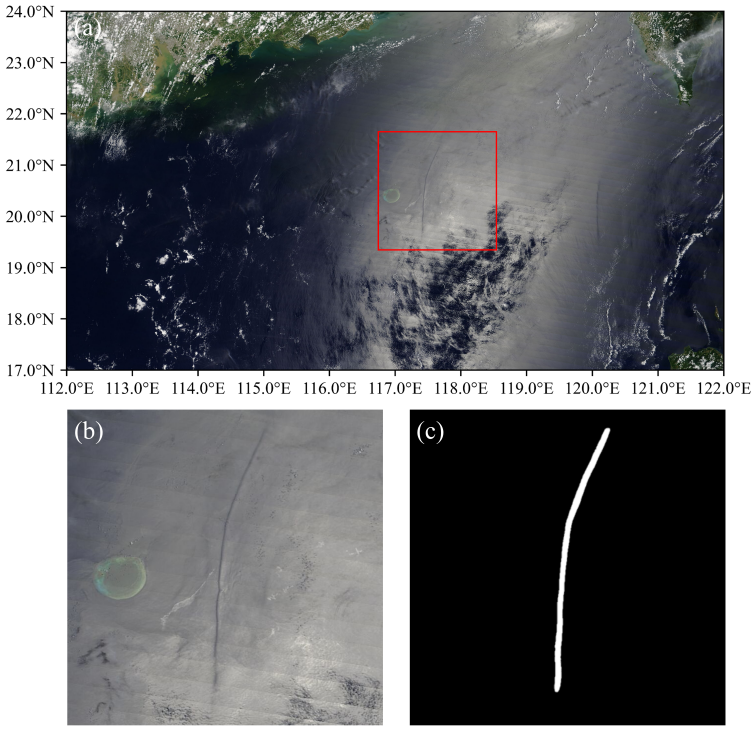

Figure 1. True-color MODIS corrected reflectance image and corresponding ISW annotation from the northern South China Sea on 19 June 2020. (**a**) MODIS true-color imagery of the study area; (**b**) 640 × 640 pixel image patch containing ISWs, corresponding to the red box in (**a**); (**c**) visualised annotation, where white pixels denote ISWs and black pixels represent the background.

The original imagery covering the study area spans a broad geographical extent and contains a large volume of data within each scene. To improve training efficiency and detection accuracy, the full-scene images were cropped into local image patches of 640 × 640 pixels while preserving the main crest structures of ISWs. Specifically, cropping and sample selection were performed using 718 original MODIS images acquired from 2000 to 2024. A non-overlapping local cropping strategy was adopted, whereby 640 × 640 pixel patches containing ISW targets were extracted from each image, ensuring that only one image patch was retained for each target region.

To construct a dataset suitable for deep learning training, the 640 × 640 pixel ISW image patches were manually annotated through visual interpretation. Only samples in which ISW features could be clearly identified were retained. These samples were required to exhibit continuous crest structures without extensive cloud contamination. Images affected by severe cloud cover or those in which ISW features could not be reliably distinguished were excluded from the dataset. The annotation process was conducted using the X-AnyLabeling tool. Its polygon annotation function was employed to delineate the contours of ISWs along the boundaries between adjacent bright and dark stripes, thereby preserving their characteristic elongated and curved morphologies. After annotation, the results were directly exported as text files in the standard YOLO format, containing the object class labels and the normalized coordinates of contour vertices for each target within the image. For the satellite image shown in Figure 1b, the corresponding annotation file was visualized using OpenCV to generate the binary image presented in Figure 1c, where white pixels represent ISWs and black pixels represent the background. Finally, 858 valid samples with dimensions of 640 × 640 pixels were obtained, each consisting of a satellite image and its corresponding annotation file. The dataset was randomly divided into training, validation, and test sets at a ratio of 8:1:1, comprising 686, 86, and 86 samples, respectively.

2.2. Model Architecture

Observation of the dataset revealed that ISW crest lines in MODIS imagery generally exhibit elongated and sinuous stripe-like structures with gradually varying boundaries. Owing to the subtle intensity differences between the crest lines and the surrounding seawater, conventional segmentation methods often struggle to achieve accurate delineation. YOLO (You Only Look Once) is a pioneering single-stage fully convolutional network originally designed for real-time object detection tasks [27]. Built upon the YOLO architecture [28], YOLO11 extends beyond conventional bounding-box detection to support efficient real-time instance segmentation [29].

Therefore, YOLO11 provides a suitable foundation for ISW detection and segmentation in satellite imagery. By integrating detection, segmentation, and parameter extraction within a unified framework, the proposed YOLO11-ISW model offers an end-to-end approach for automated ISW detection, segmentation, and geographic localization from remote sensing imagery. To better accommodate the morphological characteristics of ISWs, DSConv and the CBAM were incorporated into YOLO11, thereby further improving the robustness and accuracy of ISW detection and segmentation.

The overall architecture of the proposed YOLO11-ISW model is illustrated in Figure 2, comprising three main components: the Backbone (core feature extraction network), the Neck (feature fusion network), and the Head (detection segmentation head). The Backbone progressively extracts features from the input MODIS imagery, capturing low-level representations such as edges and textures in the early stages and generating highly discriminative high-level semantic features through deeper layers. The Neck aggregates multi-scale feature maps to integrate local details with global semantic information. Based on the fused features, the Head simultaneously performs ISW object detection and pixel-

level segmentation, producing bounding boxes, confidence scores, and crest line masks. In addition, the architectures of the SPPF [30], C3K2 [31], and C2PSA [32] modules are presented in Figure 3. These modules are designed to enlarge the receptive field, facilitate multi-scale feature fusion, and enhance feature representation within the backbone network, respectively, thereby further improving the model's feature extraction and fusion capabilities.

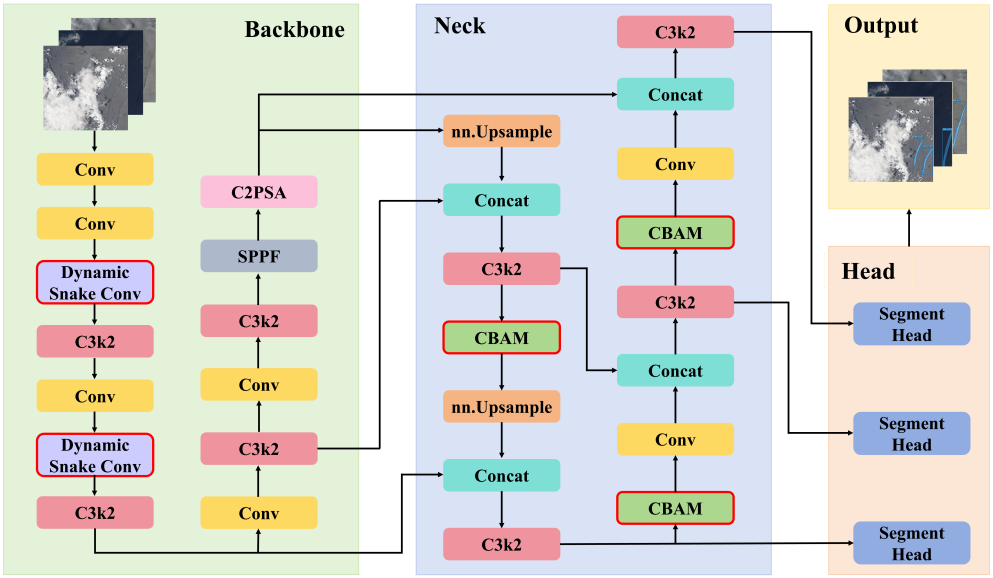

Figure 2. Overall architecture of the proposed YOLO11-ISW model. DSConv modules are incorporated into the Backbone for curved ISW feature representation, while CBAMs are inserted into the Neck for attention-based feature refinement.

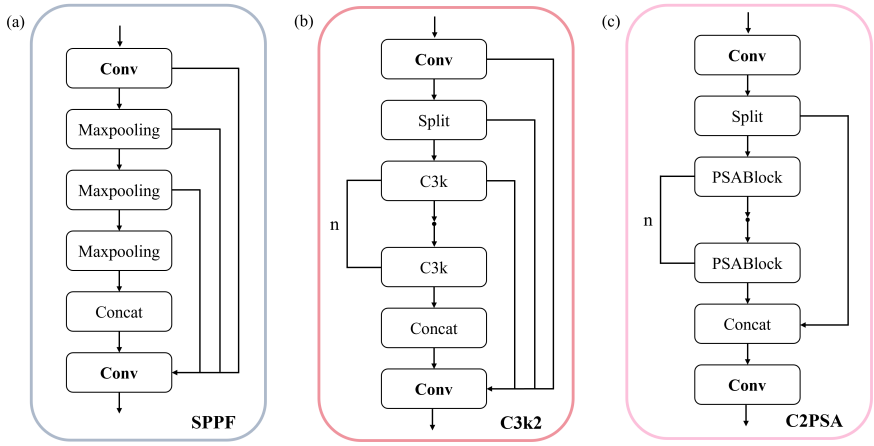

Figure 3. Architectures of the (**a**) SPPF, (**b**) C3K2, and (**c**) C2PSA modules.

2.3. Model Design Rationale

In satellite imagery, ISW crest lines typically appear as elongated and sinuous structures. Conventional convolution operations employ fixed sampling locations arranged on regular rectangular grids, making it difficult to adapt to such curved morphologies. As a result, a considerable proportion of computational resources is devoted to background regions, leading to inefficient extraction of crest line features. DSConv is a convolutional operation specifically designed for elongated, curved, and continuous structures, such as blood vessels and roads [33]. The architecture of the DSConv module is illustrated in Figure 4 [33]. By dynamically learning the sampling positions of convolution kernels, DSConv

enables adaptive deformation of the kernels along the orientation of the target, thereby facilitating more accurate extraction of structural and local features.

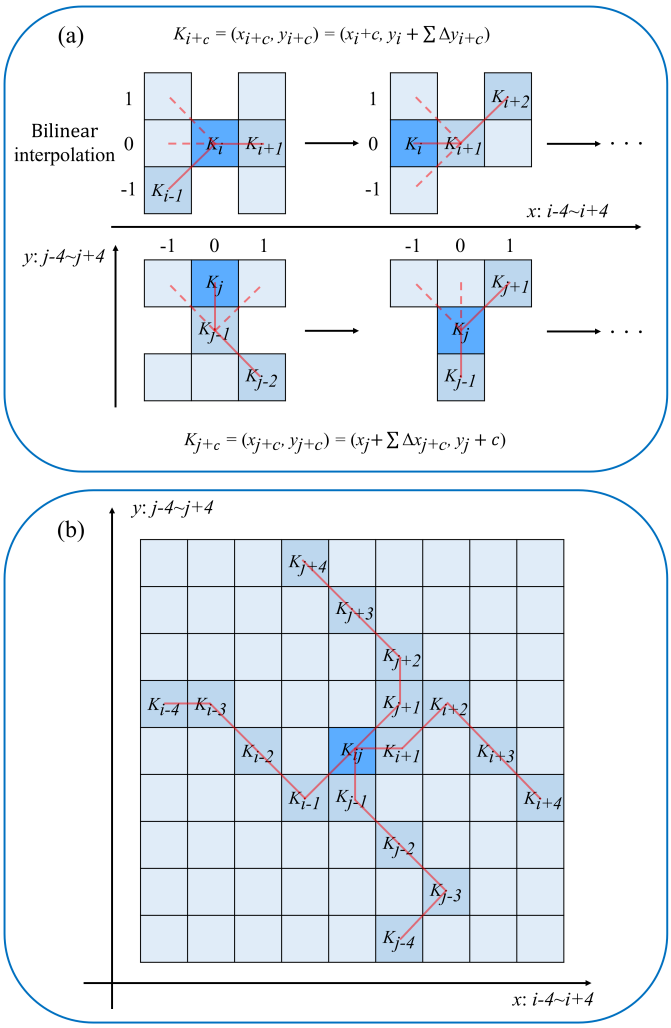

Figure 4. Dynamic snake convolution module architecture. (**a**) Schematic diagram of coordinate calculation; (**b**) the adaptive receptive field.

Therefore, DSConv was incorporated into the feature extraction layers of the YOLO11 Backbone. During forward propagation, image features first undergo preliminary extraction through the original convolution operations and are subsequently fed into the DSConv module for adaptive sampling and feature enhancement. The refined feature maps are then propagated to subsequent network layers for multi-level feature encoding. This integration strategy fully exploits the hierarchical feature extraction capability of the backbone network while specifically strengthening the model's ability to represent the continuous and curved morphologies of ISW crest lines, ultimately improving the accuracy and robustness of ISW detection and segmentation.

On the other hand, ISWs occupy only a small proportion of pixels in remote sensing imagery and often appear as weak signals with low contrast against the surrounding ocean background. The CBAM is a lightweight attention mechanism, and its architecture is illustrated in Figure 5. Through the complementary effects of channel and spatial attention, CBAM enables the network to adaptively emphasise informative features relevant to the target [34]. This joint attention mechanism is well suited to highlighting weak and spatially sparse ISW features in satellite imagery. Specifically, the channel attention module generates channel-wise weights using average pooling and max pooling operations, thereby

highlighting feature channels associated with ISWs. Building upon this process, the spatial attention module further computes spatial weights to emphasise the local regions where ISWs are located.

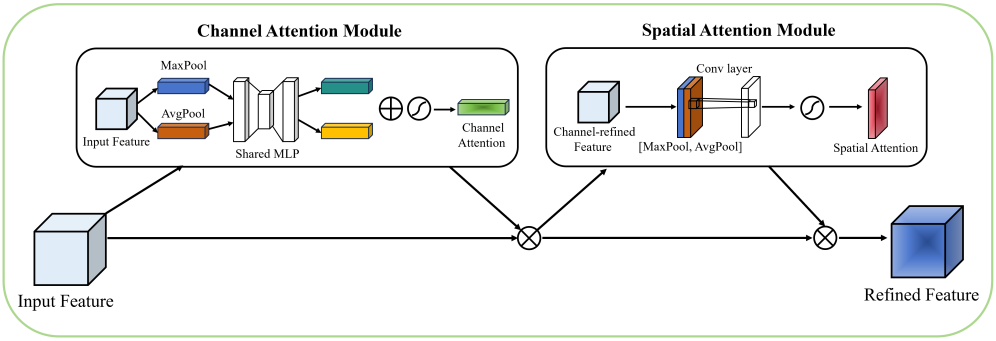

Figure 5. CBAM architecture.

The Neck is responsible for multi-scale feature fusion, during which features from different levels are combined while inevitably incorporating substantial background information. Consequently, the low-contrast characteristics of ISWs may be diluted during the fusion process. To address this issue, CBAM was embedded within the Neck for feature refinement. After feature maps from different scales are concatenated and fused, they are first passed through the CBAM for attention weighting. Channel attention is applied to select informative feature channels, followed by spatial attention to enhance target-related regions. The attention-refined features are then propagated to subsequent network layers for further processing. This integration strategy fully exploits the feature aggregation capability of the neck while guiding the network to focus on informative features through the joint effects of channel and spatial attention, thereby improving the detection sensitivity and segmentation accuracy for low-contrast ISWs.

In summary, DSConv is designed to accurately capture the local geometric characteristics of elongated and sinuous targets, whereas the CBAM attention mechanism focuses on adaptive feature reweighting to enhance the weak, low-contrast characteristics of ISWs. These two modules complement each other by enhancing local structural representation and global feature selection, thereby improving the network's ability to capture complex ISW features. Their integration constitutes a multi-level enhancement strategy tailored for ISW detection and segmentation.

2.4. Loss Function

The overall loss of the YOLO11-ISW model is formulated as a weighted combination of classification loss, bounding box regression loss, and segmentation loss, expressed as:

$$L_{ISW} = \lambda_{cls} L_{cls} + \lambda_{box} L_{box} + \lambda_{seg} L_{seg}. \quad (1)$$

To address the low proportion of ISW pixels and the imbalance between positive and negative samples in satellite imagery, focal loss [35] was adopted as the classification loss. By assigning greater emphasis to hard samples and sparse target features, focal loss effectively alleviates the impact of class imbalance. For bounding-box regression, Complete Intersection over Union (CIoU) loss [36] was employed. By jointly considering overlap area, center-point distance, and aspect-ratio consistency, CIoU loss enhances localization accuracy. For the segmentation task, a hybrid loss combining Dice loss [37] and IoU loss [38] was utilised. Specifically, Dice loss is used to alleviate the severe foreground–background imbalance caused by the relatively small proportion of ISW pixels in satellite imagery while improving the overlap and completeness of the predicted ISW regions [37,39]. IoU loss

provides an additional constraint on the overall spatial overlap between the predicted and ground-truth regions, helping to reduce incomplete or excessively expanded segmentation results [38]. Therefore, the combination of Dice and IoU losses jointly considers class imbalance, regional overlap, and spatial consistency, making it particularly suitable for the elongated, curved, and low-contrast morphology of ISWs.

3. Results

3.1. Experimental Setup

The ISW dataset described in Section 2 was employed for model training and evaluation. All experiments were conducted using an NVIDIA Tesla A100 GPU with 80 GB of memory. The YOLO11-ISW model was implemented based on the PyTorch deep learning framework [40]. The experimental environment was configured with CUDA 13.0, Python 3.12, and PyTorch 2.9.0. The principal training hyperparameters were set as follows: an initial learning rate of 0.005, a batch size of 16, and 500 training epochs. The AdamW optimizer [41] was adopted during training. Compared with the conventional Adam optimizer, AdamW decouples weight decay from gradient updates, thereby preventing the regularisation effect of weight decay from being weakened by adaptive learning rates. This design contributes to mitigating overfitting and is well suited to the sparse characteristics of ISWs in satellite imagery.

The training loss curves of the YOLO11-ISW model are presented in Figure 6. All three loss components exhibit a pattern of rapid reduction during the early stages of training, followed by gradual convergence and eventual stabilization, indicating a stable and effective optimization process. Specifically, the classification loss converges to the lowest final value, suggesting that Focal Loss effectively alleviates the imbalance between ISW and background samples and enables robust class discrimination. The stable convergence of the bounding box regression loss indicates that CIoU Loss facilitates accurate localization of ISWs, thereby providing reliable spatial constraints for the subsequent segmentation task. Although the segmentation loss converges to a relatively higher value, it consistently decreases throughout training, demonstrating the effectiveness of the hybrid segmentation loss in handling sparse ISW pixels and diffuse boundary characteristics. Overall, the loss curves indicate that the model has successfully learned the classification, localization, and segmentation characteristics of ISWs, resulting in favourable detection and segmentation performance.

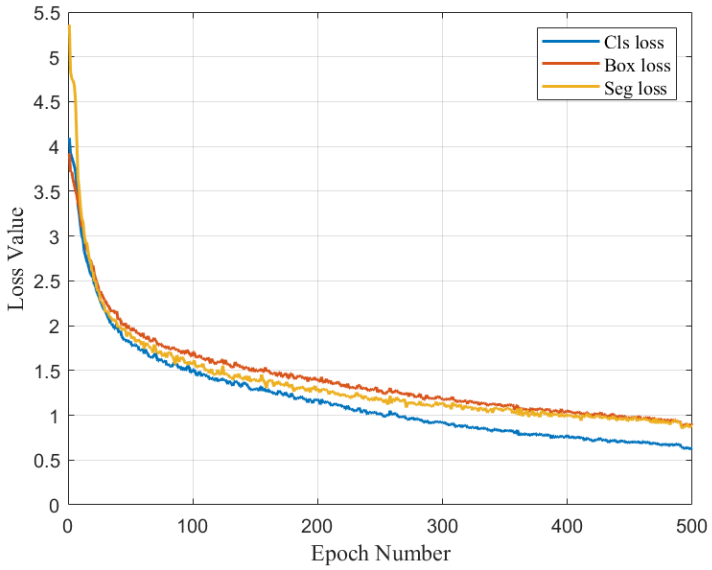

Figure 6. Training loss curves of the proposed YOLO11-ISW model.

3.2. Evaluation Metrics

To comprehensively and objectively evaluate the performance of the YOLO11-ISW model in ISW detection and segmentation tasks, Precision, Recall, and F1-score were adopted as quantitative evaluation metrics.

Precision represents the proportion of correctly identified ISW samples among all samples predicted as ISWs by the model. It reflects the reliability of the model predictions and is defined as follows [42]:

$$Precision = \frac{TP}{TP + FP}. \tag{2}$$

Recall represents the proportion of true ISW samples that are successfully identified by the model among all actual ISW samples. It reflects the model's capability to detect positive targets and is defined as follows [43]:

$$Recall = \frac{TP}{TP + FN}. \tag{3}$$

F1-score is the harmonic mean of Precision and Recall and provides a comprehensive assessment of the model performance on the positive class. It is calculated as follows [44]:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FN + FP}. \tag{4}$$

Figures 7 and 8 illustrate the variations in Precision, Recall, and F1-score during the training process of the YOLO11-ISW model for ISW crest line detection and segmentation, respectively. Overall, the model demonstrates rapid learning, efficient convergence, and stable performance across both tasks, indicating its effectiveness for ISW identification.

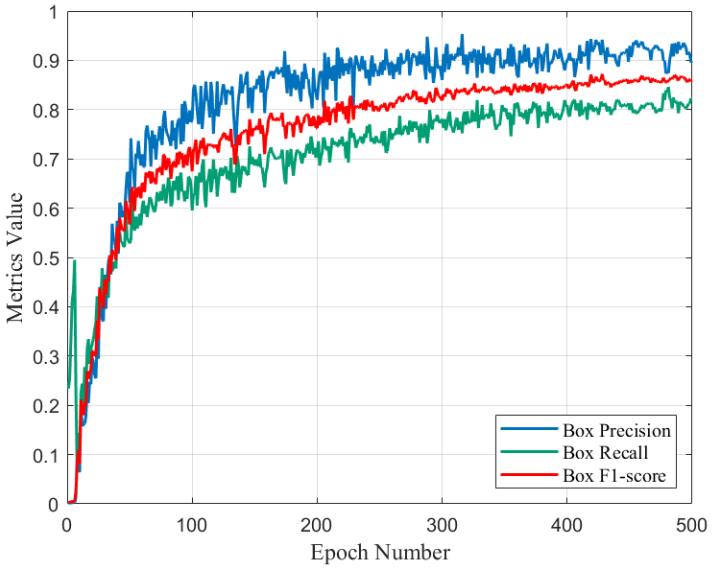

Figure 7. Precision, Recall, and F1-score for ISW detection during training.

Specifically, the model exhibits a fast learning stage at the early phase of training, where Precision, Recall, and F1-score increase rapidly within the first 100 iterations. This suggests that the model can quickly capture the basic visual features and spatial distribution of ISWs in satellite imagery. From 100 to 300 iterations, all metrics continue to improve steadily and gradually stabilise around 300 iterations, indicating that the model has reached convergence without evident overfitting. After convergence, for the detection task, Precision remains around 0.90, indicating a low false-positive rate, while Recall stabilises at approximately

0.80, demonstrating the model's capability to detect scattered ISW targets with varying scales. The F1-score approaches 0.85, reflecting a balanced trade-off between Precision and Recall. A similar trend is observed for the segmentation task, where the model effectively distinguishes ISW boundaries from the surrounding sea surface, ensuring the integrity of segmentation masks. Comparison of the curves of the two tasks shows that both achieve high precision, while recall remains relatively low. This is mainly attributed to the presence of low-contrast and weak-boundary ISWs, which may lead to occasional missed detections. In addition, the segmentation task requires reconstruction of complex sinuous crest structures, making it inherently more challenging and resulting in slightly lower Recall than Precision. Overall, the smooth evolution of all metrics indicates a stable and well-balanced model performance.

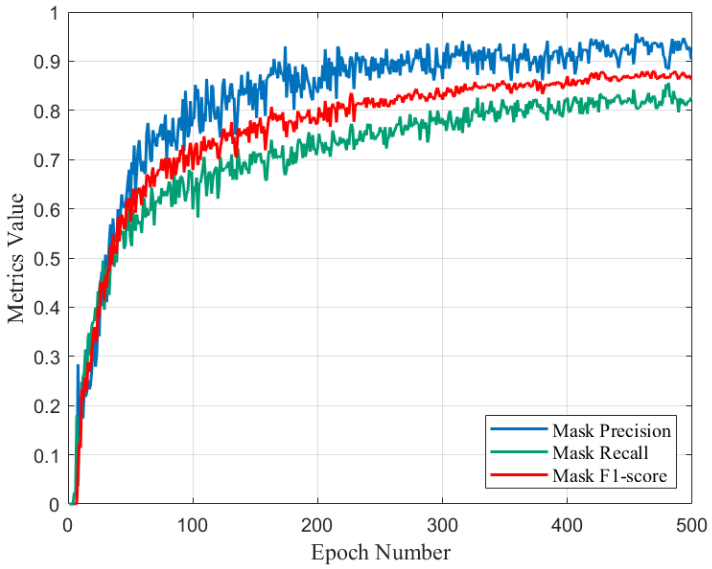

Figure 8. Precision, Recall, and F1-score for ISW segmentation during training.

3.3. Ablation Experiments

To address the major challenges of low contrast and elongated, sinuous morphology in ISW crest line recognition, DSConv and the CBAM attention mechanism were incorporated into the YOLO11 framework to enhance morphological modeling and weak feature extraction, respectively. To validate the effectiveness of each module and their synergistic contributions, ablation experiments were conducted using a controlled experimental design. The results are presented in Table 1.

Table 1. Ablation experiment results.

Metrics		Models			
		YOLO11	**YOLO11 + DSConv**	**YOLO11 + CBAM**	**YOLO11-ISW**
Params		2.82 M	2.86 M	2.86 M	2.90 M
GFLOPS		10.8 G	10.9 G	10.8 G	10.9 G
Box	**Precision**	88.61%	91.45%	89.48%	92.01%
	Recall	78.09%	78.74%	78.93%	80.73%
	F1-score	83.02%	84.62%	83.87%	86.00%
Mask	**Precision**	89.41%	92.24%	89.52%	93.14%
	Recall	79.06%	79.41%	79.89%	82.60%
	F1-score	83.92%	85.35%	84.43%	87.55%

Compared with the original YOLO11 model, introducing DSConv alone increased the detection F1-score from 83.02% to 84.62%. This improvement indicates that DSConv can better capture the elongated and curved structural characteristics of ISW crest lines, thereby enhancing the representation of irregular target contours. In contrast, incorporating CBAM alone improved the segmentation F1-score from 83.92% to 84.43%. This result suggests that the channel and spatial attention mechanism can enhance informative features and highlight the spatial regions associated with weak-contrast ISW signals, which is beneficial for pixel-level segmentation. When DSConv and CBAM were integrated simultaneously, the YOLO11-ISW model achieved the best overall performance, indicating that the two modules provide complementary benefits. DSConv primarily enhances the representation of the elongated and curved morphology of ISWs, whereas CBAM further refines the extracted features by emphasising informative responses and relevant spatial regions. Therefore, their combination enables the network to simultaneously improve morphological representation and weak-feature discrimination.

For the detection task, YOLO11-ISW achieved a Precision of 92.01%, a Recall of 80.73%, and an F1-score of 86.00%. For the segmentation task, the corresponding values were 93.14%, 82.60%, and 87.55%, respectively. Compared with the baseline model, these improvements were achieved with only marginal increases in the number of parameters and computational cost. Overall, the ablation results demonstrate that the proposed modifications improve the network in a task-oriented manner while preserving its lightweight characteristics.

3.4. Model Performance

To further evaluate the effectiveness of the proposed YOLO11-ISW model, several representative object detection and semantic segmentation approaches, including Faster R-CNN [45], YOLOv7 [46], SegFormer [47], PAU-Net [19], TransUNet [48], and IWE-Net [**?**], were reproduced and implemented for comparative experiments. All models were trained and evaluated using the same dataset and experimental configurations to provide a fair comparison. In addition to detection and segmentation performance, the inference time of each model was evaluated to assess its computational efficiency. The inference time represents the time required to process a single 640 × 640 satellite image. The corresponding results are presented in Table 2.

Table 2. Performance comparison of YOLO11-ISW with representative detection and segmentation methods.

Method	Box			Mask			Inference Speed
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
Faster R-CNN	81.14%	70.65%	75.53%	–	–	–	23.63 ms
YOLOv7	86.98%	61.63%	72.14%	–	–	–	7.40 ms
SegFormer	–	–	–	75.28%	78.87%	77.03%	11.93 ms
PAU-Net	–	–	–	72.47%	76.82%	74.58%	4.48 ms
TransUnet	–	–	–	76.70%	70.62%	73.53%	7.74 ms
IWE-Net	–	–	–	78.03%	87.46%	81.56%	8.05 ms
YOLO11-ISW	92.01%	80.73%	86.00%	93.14%	82.60%	87.55%	10.97 ms

As a classical two-stage detection algorithm, Faster R-CNN provides relatively stable localization performance but suffers from slow inference speed and cannot directly generate segmentation results. YOLOv7 exhibits favourable real-time detection capability, achieving an inference time of only 7.40 ms. However, it lacks pixel-level segmentation capability and shows relatively low recall for low-contrast ISWs. Among the segmentation models, SegFormer benefits from the Transformer architecture and demonstrates strong global contextual modeling capability, but it suffers from relatively slow inference speed. PAU-

Net and TransUNet achieve slightly faster inference, but their segmentation performance remains inferior to that of YOLO11-ISW, and neither model is capable of performing detection and segmentation simultaneously. Compared with IWE-Net, YOLO11-ISW achieved higher segmentation Precision and F1-score, although with a lower Recall. This indicates that YOLO11-ISW provides more precise segmentation, whereas IWE-Net achieves higher pixel-level recall. More importantly, IWE-Net primarily focuses on pixel-level binary extraction, whereas YOLO11-ISW jointly performs object detection and pixel-level segmentation, providing both explicit object-level localization and detailed spatial delineation of ISWs.

In contrast to these existing approaches, YOLO11-ISW enables end-to-end detection and segmentation within a lightweight architecture comprising only 2.90 M parameters. The detection F1-score reaches 86.00%, outperforming that of Faster R-CNN and YOLOv7, while the segmentation F1-score reaches 87.55%, exceeding that of SegFormer, PAU-Net, TransUNet, and IWE-Net. In addition, YOLO11-ISW achieves an inference time of 10.97 ms, demonstrating a favourable balance among accuracy, efficiency, and model compactness.

To provide an intuitive comparison of model performance, the segmentation results obtained on the test set were visualized, as shown in Figure 9. The segmentation outputs produced by different models are presented for comparison.

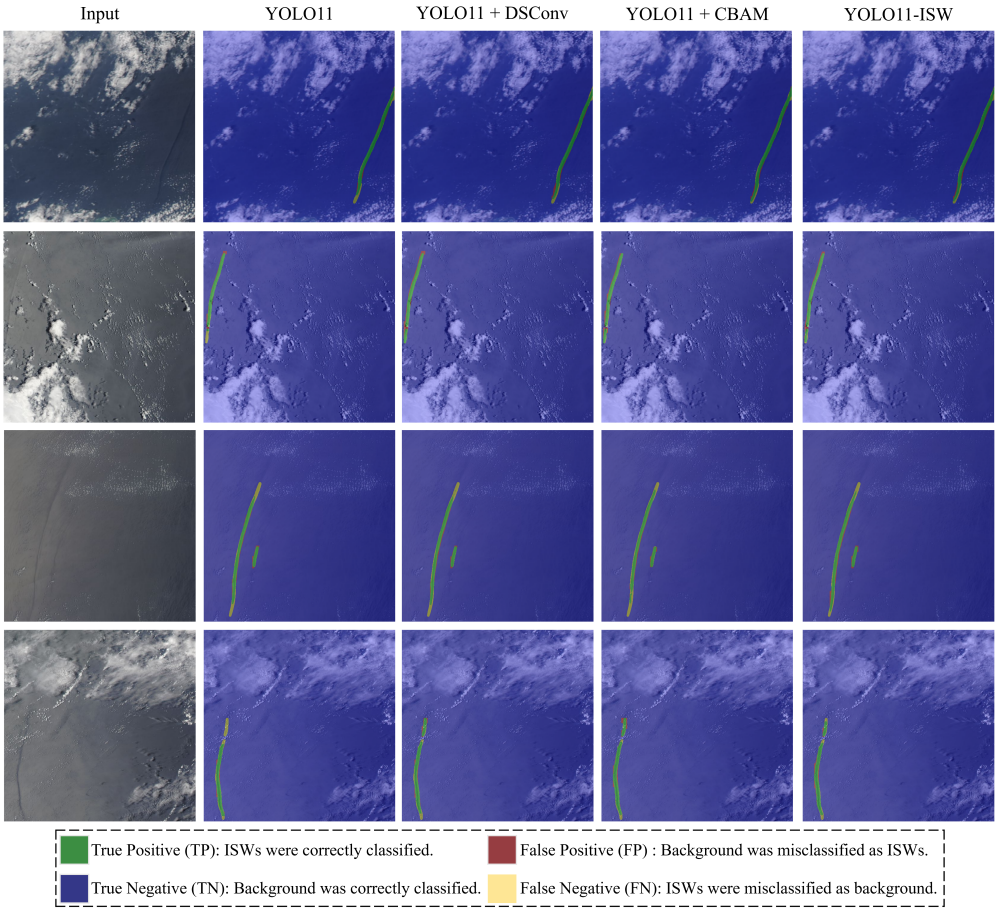

Figure 9. Comparison of segmentation results across different models.

The original YOLO11 model (second column) exhibits a relatively large number of yellow pixels, which represent false negatives, indicating missed detections along the sinuous crest structures of ISWs. After incorporating DSConv (third column), the previously missed waveform regions are successfully identified, and the overall predicted contours show closer agreement with the actual morphology of ISWs. This observation demonstrates that DSConv effectively enhances the model's capability to capture elongated and curved

waveform features. Following the introduction of the CBAM (fourth column), the number of yellow pixels is substantially reduced, indicating fewer false negatives and further alleviation of the missed-detection problem. This finding confirms the effectiveness of CBAM in strengthening the representation of low-contrast ISW features.

The proposed YOLO11-ISW model (fifth column) produces the fewest yellow and red pixels, representing the lowest rates of false negatives and false positives, respectively. The extracted waveforms are the most complete, while both missed detections and false alarms are substantially reduced. These results demonstrate the complementary advantages of DSConv and CBAM, highlighting the synergistic improvement achieved through their combined application.

The proposed YOLO11-ISW model enables accurate detection and segmentation of ISWs in satellite imagery. Building upon these results, the embedded geospatial information contained in GeoTIFF data was utilised to transform the pixel-level segmentation outputs into geographical coordinates, thereby obtaining the actual latitude and longitude information of ISWs. These geolocation data provide reliable support for quantitative studies of ISWs, analyses of their spatial distribution patterns, and the safety assessment of marine engineering activities.

The geographical coordinates derived from the identified ISWs are illustrated in Figure 10. Figure 10a presents the original satellite image used as the model input. Figure 10b shows the detection and segmentation results produced by the YOLO11-ISW model, where the blue bounding boxes indicate the detected ISWs and the blue contours represent their segmentation masks. The converted geographical coordinates were subsequently visualized, as shown in Figure 10c. The blue outlines represent the geospatial distribution of the segmented ISWs, while the red lines denote the extracted crest lines, clearly revealing their spatial morphology.

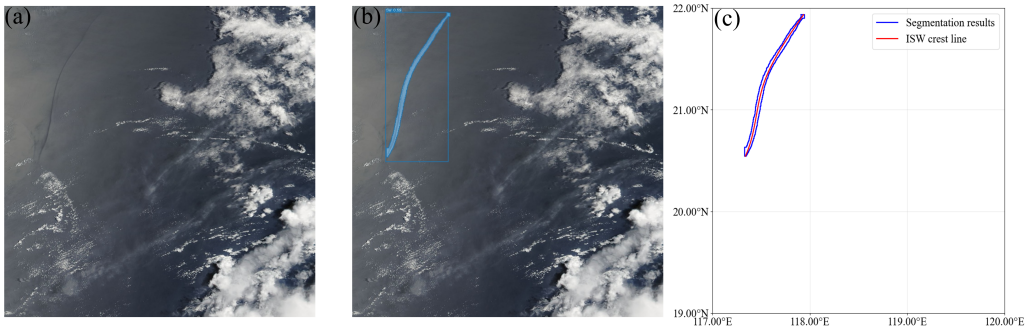

Figure 10. Georeferenced ISW detection, segmentation, and crest-line extraction results: (**a**) original satellite image; (**b**) ISW detection and segmentation results produced by the YOLO11-ISW model; and (**c**) georeferenced ISW segmentation results and extracted crest line.

3.5. Spatiotemporal Distribution of ISWs in the Northern South China Sea

To further evaluate the applicability of the proposed YOLO11-ISW framework to long-term and large-scale observations, Terra and Aqua MODIS imagery acquired over the northern South China Sea from 2011 to 2020 was analyzed. A total of 7306 satellite images were collected, comprising 3653 Terra and 3653 Aqua observations. Using the sliding-window inference strategy, the YOLO11-ISW model automatically identified and extracted ISW signatures over the entire study area. ISWs were detected in 1288 images, corresponding to 17.63% of all analyzed observations. It should be emphasized that the statistics reported here represent satellite-observed ISW detections rather than the absolute occurrence frequency of ISWs in the ocean.

The spatial distribution derived from the 2011–2020 observations exhibits pronounced regional heterogeneity (Figure 11). The most prominent feature is a relatively continuous high-detection-frequency band located approximately between 115.5°–117.5°E and 20.8°–21.8°N. Rather than being distributed uniformly over the northern South China Sea, the detected ISWs are organized into distinct spatial clusters. This pattern is consistent with previous long-term satellite observations. Based on 22 years of MODIS imagery, Zhang and Li [49] similarly reported a clustered distribution of internal waves in the northern South China Sea, with a particularly high concentration in the broad region surrounding Dongsha Atoll and the continental-slope transition zone. Earlier satellite observations also documented frequent westward-propagating ISWs across the northern South China Sea and toward the continental shelf [50]. Thus, the high-frequency band identified in the present study is not merely a feature arising from the 2011–2020 observation period, but agrees well with the major regions of enhanced ISW activity repeatedly reported in previous satellite-based studies.

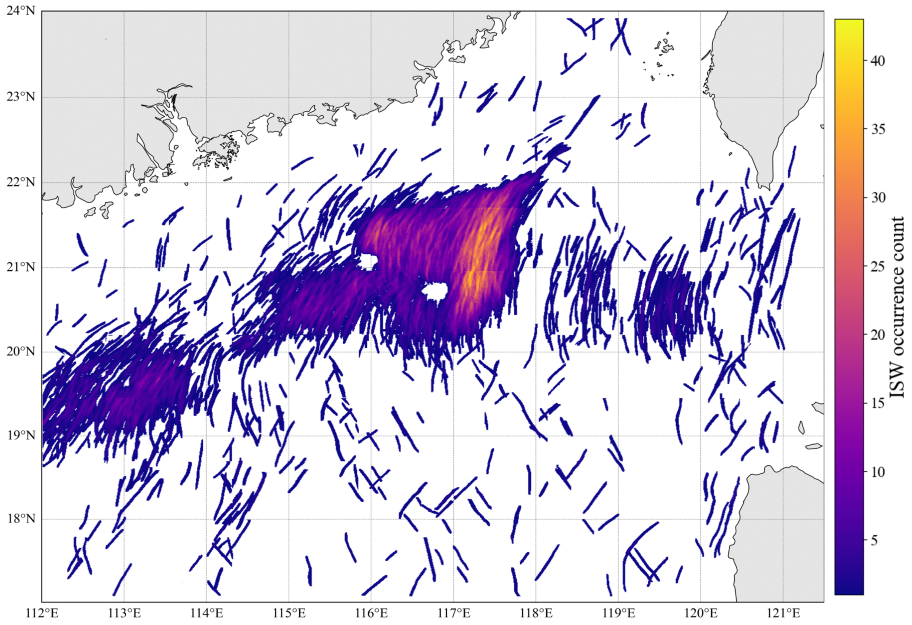

Figure 11. Spatial distribution of ISW occurrences detected from Terra and Aqua satellite imagery in the northern South China Sea during 2011–2020.

The spatial pattern also reveals several secondary high-detection-frequency regions in the southwestern, eastern, and northeastern parts of the study area, together with two pronounced low-detection-frequency regions located around Dongsha Atoll and to its northwest. Similar spatial heterogeneity was documented by Zhang and Li [49], who identified persistent low-detection-frequency regions in the long-term satellite record near Dongsha Atoll, which were linked to the influence of surrounding complex bathymetry. The spatial distribution of these low-detection regions may result from the complex interaction between propagating ISWs and local topographic features. Previous satellite observations have demonstrated that ISW crest lines undergo substantial refraction, deformation, and spatial reorganization when passing around Dongsha Atoll [51]. Such topographic interactions can modify the geometry and surface expression of ISWs, potentially affecting the continuity and detectability of wave signatures in satellite imagery. Overall, the spatial distribution obtained by YOLO11-ISW is consistent with the major features reported in previous remote-sensing studies [49,51], supporting the capability of the proposed framework to characterize large-scale satellite-observed ISW distribution patterns from long-term remote-sensing observations.

The temporal variability of satellite-observed ISW detections was further examined at interannual, seasonal, monthly, and lunar-day scales (Figure 12). At the interannual scale, the annual number of satellite images containing detected ISW signatures varied between 111 and 153 during 2011-2020 (Figure 12a). Although year-to-year fluctuations are evident, no persistent increasing or decreasing trend was observed over the decade-long observation period. Previous satellite-based analyses have similarly reported substantial interannual variability in satellite-observed ISW occurrence in the northern South China Sea [8], while long-term mooring observations have demonstrated considerable year-to-year variations in ISW intensity [52]. These findings suggest that interannual variability represents an important characteristic of ISW variability in this region.

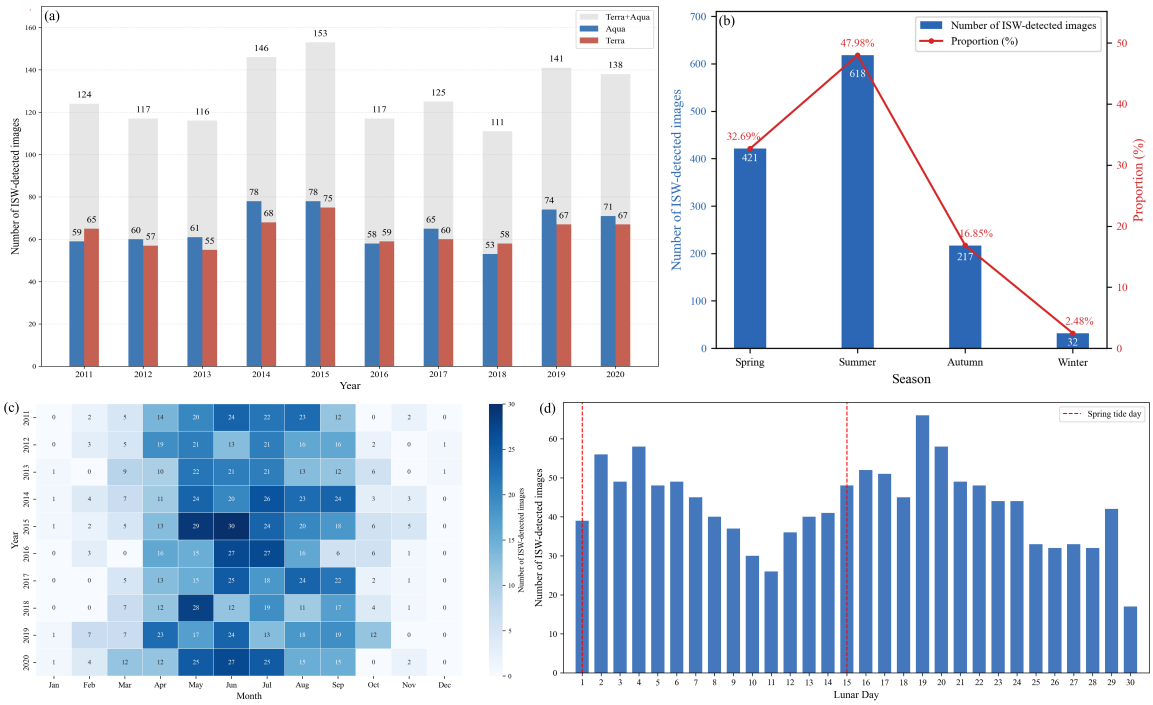

Figure 12. Temporal distribution of satellite-observed ISW detections in the northern South China Sea during 2011–2020 at different temporal scales: (**a**) interannual variation; (**b**) seasonal distribution; (**c**) monthly distribution; and (**d**) lunar-day distribution.

A clear seasonal pattern emerges in the distribution of satellite-observed ISW detections (Figure 12b). Summer accounts for the largest proportion of ISW-detected images (48.0%), followed by spring (32.7%) and autumn (16.8%), whereas only 2.5% occur in winter. Thus, more than 80% of all satellite-observed ISW detections are concentrated in spring and summer. This pronounced seasonality agrees closely with previous studies of the northern South China Sea [8,49]. The consistency among different sensors, observational periods, and analysis methods indicates that the strong spring–summer concentration identified by YOLO11-ISW represents a robust seasonal feature of satellite-observed ISWs in the northern South China Sea. The observed seasonal variability may be related to seasonal changes in the upper-ocean environment, which influence ISW generation, propagation, and surface manifestation. Cai et al. [53,54] showed that ISW nonlinearity in the South China Sea is sensitive to stratification and is generally stronger in summer than in winter.

The monthly distribution provides a more detailed view of the seasonal variability in satellite-observed ISW detections (Figure 12c). Detection counts begin to increase rapidly in March, remain high during May–July, and reach their maximum in June before gradually decreasing through autumn and winter. Particularly noteworthy is the June maximum, which closely reproduces the satellite climatology reported by Zheng et al. [8],

who found the highest SAR-observed occurrence frequency in June and markedly lower frequencies from December to February. Previous observations by Cai et al. [53] also consistently indicate a maximum ISW occurrence frequency around June–July and a minimum during January–February. The agreement between the present automated MODIS analysis and these earlier manually interpreted or independently derived statistics suggests that YOLO11-ISW is capable not only of detecting individual ISW signatures but also of capturing established climatological temporal patterns when applied to a decade-long satellite archive.

Because ISW variability in the northern South China Sea has been shown to exhibit a pronounced fortnightly modulation associated with the spring–neap tidal cycle [52], the distribution of satellite-observed ISW detections with respect to lunar day was further examined. Satellite-observed ISW detections exhibit a distinct non-uniform distribution with respect to lunar day (Figure 12d). Two enhanced detection periods occur around the spring-tide phases, producing an approximately fortnightly double-peak pattern. Importantly, the maxima do not coincide exactly with the theoretical spring-tide days but generally appear several days later. Zhang and Li [49] identified a similar double-peak distribution from their 22-year MODIS dataset and reported that the maximum number of observed internal waves occurred approximately four days after spring tide. They suggested that this delayed response was mainly related to the propagation time required for internal waves generated near the Luzon Strait to reach the continental-shelf regions. The observed lag in this study may therefore be associated with the combined effects of tidal forcing, nonlinear wave evolution, and westward propagation processes.

4. Discussion

The proposed YOLO11-ISW model provides an effective approach for automated detection and pixel-level segmentation of ISWs from satellite remote sensing imagery. Compared with conventional manual interpretation methods, the proposed deep learning framework significantly improves the efficiency and consistency of large-scale ISW identification. The improved performance can be attributed to the integration of DSConv and CBAM, which enhance the extraction of weak and elongated ISW-related features while preserving complex crest-line structures. Furthermore, the combination of object detection and instance segmentation enables not only the localization of ISWs but also the extraction of detailed spatial information, providing a reliable basis for subsequent quantitative analysis. The long-term MODIS observations further demonstrate that the proposed framework can be applied for regional-scale ISW monitoring. The detected spatial and temporal distribution patterns are generally consistent with previous studies in the northern South China Sea, confirming the reliability of the model-derived observations. The concentration of detected ISWs west of the Luzon Strait agrees with previous findings that this region is the primary generation area of energetic internal tides, which subsequently evolve into nonlinear internal solitary waves during their westward propagation. The seasonal variability identified in this study, with enhanced ISW detections during spring and summer and reduced detections during winter, is also consistent with previous investigations and may be related to seasonal variations in upper-ocean stratification and thermocline structure. In addition, the observed relationship between ISW detection frequency and lunar tidal cycles further highlights the dominant role of tidal forcing in ISW generation. The delayed peak relative to spring tide is consistent with previous satellite-based studies, suggesting that the propagation time from the Luzon Strait and subsequent nonlinear evolution processes contribute to the lag between tidal forcing and satellite-observed ISW signatures.

Although the present study focuses on automatic detection, segmentation, and spatial–temporal characterization of ISWs, the extracted crest-line information provides a potential

foundation for further dynamic analysis. By combining georeferenced crest lines from sequential satellite observations, the displacement of ISWs can theoretically be derived, allowing the estimation of propagation direction and velocity. Such an extension would provide an opportunity to investigate ISW kinematics over broad spatial scales based on remote sensing observations. However, reliable propagation estimation requires appropriate multi-temporal observations with sufficient temporal resolution. The revisit interval of polar-orbiting satellites such as MODIS may limit continuous tracking of rapidly evolving ISWs, particularly for waves with relatively short propagation times or complex evolution processes. Therefore, while the current framework establishes the capability of extracting accurate ISW spatial information, further development of automated propagation analysis requires the integration of higher-temporal-resolution observations and additional validation using independent measurements.

Despite the promising performance of the proposed framework, several limitations should be considered. First, the current model was developed and evaluated using optical satellite imagery from the northern South China Sea. Although the model achieved satisfactory performance within the study region, its applicability to other oceanic environments with different hydrographic conditions, background variability, and ISW morphological characteristics requires further investigation. Second, optical remote sensing observations are inevitably affected by environmental factors, including cloud coverage, solar illumination conditions, atmospheric interference, and variations in sea surface roughness, which may influence the visibility of ISW signatures and introduce uncertainties into the detection results. Future studies could improve the robustness and generalization capability of the framework by incorporating multi-source remote sensing data, such as synthetic aperture radar (SAR) and other optical sensors with different spatial and temporal resolutions. In addition, combining satellite observations with numerical ocean models and in situ measurements may provide a more comprehensive understanding of ISW generation, propagation, and transformation processes. These developments would further enhance the capability of deep learning-based remote sensing approaches for long-term and large-scale monitoring of internal wave dynamics.

5. Conclusions

ISWs are important nonlinear ocean dynamical phenomena whose spatial distribution and propagation characteristics are closely related to marine engineering safety, ocean environmental processes, and marine resource utilization. Therefore, accurate detection and segmentation of ISWs are fundamental prerequisites for related oceanographic studies. MODIS satellite imagery provides wide spatial coverage, frequent observations, and long-term data availability, making it suitable for large-scale and long-term monitoring of ISWs. To develop an efficient and accurate model for ISW detection and segmentation, the YOLO11 framework was further improved. DSConv was embedded into the Backbone to enhance the model's ability to capture the continuous morphological characteristics of ISWs by exploiting its strong representation capability for curved structures. In addition, CBAM was incorporated into the Neck to enhance informative ISW features through channel and spatial attention. These improvements enhanced the detection sensitivity and segmentation accuracy for low-contrast ISWs, leading to the development of the YOLO11-ISW model for ISW detection and segmentation.

Experimental results demonstrated that the proposed YOLO11-ISW model achieved reliable performance in ISW detection and segmentation tasks. For object detection, the Precision, Recall, and F1-score reached 92.01%, 80.73%, and 86.00%, respectively. For segmentation, the corresponding values were 93.14%, 82.60%, and 87.55%, respectively. Compared with Faster R-CNN, YOLOv7, SegFormer, PAU-Net, TransUNet, IWE-Net and the original

YOLO11 model, the proposed method exhibited superior overall performance, enabling accurate detection and segmentation of ISWs under complex ocean conditions. Furthermore, the proposed framework was applied to a decade-long MODIS dataset to investigate the spatiotemporal distribution characteristics of ISWs in the northern South China Sea. The detected ISW distribution patterns revealed significant spatial aggregation west of the Luzon Strait and pronounced seasonal and lunar-cycle variations, which were consistent with previous studies on ISW generation and propagation mechanisms in this region. These results demonstrate the potential of deep learning-based remote sensing methods for long-term and large-scale monitoring of ISW variability.

The pixel-level segmentation results and georeferenced crest-line information generated by YOLO11-ISW also provide a potential foundation for further analysis of ISW dynamics. By integrating sequential satellite observations with accurate spatial information, additional dynamic parameters, such as propagation direction and velocity, could potentially be derived in future studies.

In summary, the proposed YOLO11-ISW framework provides an effective solution for automated ISW detection, segmentation, and large-scale remote sensing monitoring. The integration of deep learning techniques with satellite observations offers new opportunities for investigating ISW spatial–temporal variability and supports future studies of internal wave dynamics, marine environmental monitoring, and ocean engineering applications.

Author Contributions: Conceptualization, G.R., Y.K. and Y.F.; methodology, G.R., H.D. and Y.F.; software, G.R.; validation, G.R., H.D. and Y.F.; formal analysis, Y.K.; data curation, Y.F.; writing—original draft preparation, G.R.; writing—review and editing, Y.F., Z.L. and G.R.; visualization, G.R.; supervision, H.D. and Y.K.; project administration, Y.F.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (Grant No. 2024YFB4206600) and the Natural Science Foundation of Shandong Province (Grant ZR2024MD091).

Data Availability Statement: The data generated and analyzed during this study are available from the corresponding author on reasonable request.

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

References

1. Jackson, C. Internal wave detection using the moderate resolution imaging spectroradiometer (MODIS). *J. Geophys. Res.-Ocean.* **2007**, *112*, C11012. https://doi.org/10.1029/2007JC004220.
2. Xu, J.; Du, H.; Peng, P.; Wang, S.; Gu, Z. Nonstationary spatiotemporal characteristics of internal solitary waves propagating along a convex side boundary. *Ocean Eng.* **2024**, *311*, 118867. https://doi.org/10.1016/j.oceaneng.2024.118867.
3. Cai, S.; Wang, S.; Long, X. A simple estimation of the force exerted by internal solitons on cylindrical piles. *Ocean Eng.* **2006**, *33*, 974–980. https://doi.org/10.1016/j.oceaneng.2005.05.012.
4. Wang, Y.H.; Dai, C.F.; Chen, Y.Y. Physical and ecological processes of internal waves on an isolated reef ecosystem in the South China Sea. *Geophys. Res. Lett.* **2007**, *34*, L18609. https://doi.org/10.1029/2007GL030658.
5. Dong, J.; Zhao, W.; Chen, H.; Meng, Z.; Shi, X.; Tian, J. Asymmetry of internal waves and its effects on the ecological environment observed in the northern South China Sea. *Deep-Sea Res. I Oceanogr. Res. Pap.* **2015**, *98*, 94–101. http://doi.org/10.1016/j.dsr.2015.01.003.
6. Ramp, S.R.; Tang, T.Y.; Duda, T.F.; Lynch, J.F.; Liu, A.K.; Chiu, C.S.; Bahr, F.L.; Kim, H.R.; Yang, Y.J. Internal solitons in the northeastern South China Sea. Part I: Sources and deep water propagation. *IEEE J. Ocean. Eng.* **2004**, *29*, 1157–1181. http://doi.org/10.1109/JOE.2004.840839.
7. Zhao, Z.; Klemas, V.; Zheng, Q.; Yan, X.H. Remote sensing evidence for baroclinic tide origin of internal solitary waves in the northeastern South China Sea. *Geophys. Res. Lett.* **2004**, *31*, L06302. http://doi.org/10.1029/2003GL019077.
8. Zheng, Q.; Susanto, R.D.; Ho, C.R.; Song, Y.T.; Xu, Q. Statistical and dynamical analyses of generation mechanisms of solitary internal waves in the northern South China Sea. *J. Geophys. Res.-Ocean.* **2007**, *112*, C03021. http://doi.org/10.1029/2006JC003551.

9. Ródenas, J.A.; Garello, R. Internal wave detection and location in SAR images using wavelet transform. *IEEE Trans. Geosci. Remote Sens.* **2002**, *36*, 1494–1507. http://doi.org/10.1109/36.718853.
10. Zheng, Y.G.; Zhang, H.S.; Wang, Y.Q. Stripe detection and recognition of oceanic internal waves from synthetic aperture radar based on support vector machine and feature fusion. *Int. J. Remote Sens.* **2021**, *42*, 6706–6724. https://doi.org/10.1080/01431161.2021.1943040.
11. Surampudi, S.; Sasanka, S. Internal wave detection and characterization with SAR data. In *Proceedings of the 2019 IEEE Recent Advances in Geoscience and Remote Sensing: Technologies, Standards and Applications (TENGARSS)*; IEEE: Kochi, India, 2019; pp. 104–108. https://doi.org/10.1109/TENGARSS48957.2019.8976045.
12. Simonin, D.; Tatnall, A.; Robinson, I. The automated detection and recognition of internal waves. *Int. J. Remote Sens.* **2009**, *30*, 4581–4598. http://doi.org/10.1080/01431160802621218.
13. Zhang, S.; Li, X.; Zhang, X. Internal wave signature extraction from SAR and optical satellite imagery based on deep learning. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4203216. http://doi.org/10.1109/TGRS.2023.3258189.
14. Ma, C.; Li, B.; Gao, Z.; Zhang, C.; Zhao, C.; Chen, G. Detecting internal waves from altimeter data using deep learning method. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4202611. http://doi.org/10.1109/TGRS.2025.3541248.
15. Bao, S.; Meng, J.; Sun, L.; Liu, Y. Detection of ocean internal waves based on Faster R-CNN in SAR images. *J. Oceanol. Limnol.* **2020**, *38*, 55–63. https://doi.org/10.1007/s00343-019-9028-6.
16. Cai, L.; Zha, G.; Lin, M.; Wang, X.; Zhang, H. Ocean internal wave detection in SAR images based on improved YOLOv7. *IEEE Access* **2024**, *12*, 146852–146865. http://doi.org/10.1109/ACCESS.2024.3468641.
17. Huo, X.; Gao, H.; Huang, B.; Chen, G. CHGAFF-YOLO: A Cascade Hybrid Global Adaptive Feature Fusion Framework for Real-Time Ocean Internal Wave Detection. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 1503205. http://doi.org/10.1109/LGRS.2025.3577597.
18. Meng, J.; Zhang, H.; Sun, L.; Wang, J. Remote sensing techniques for detecting internal solitary waves: A comprehensive review and prospects. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 46–78. http://doi.org/10.1109/MGRS.2024.3402673.
19. Ma, Y.; Meng, J.; Sun, L.; Ren, P. Oceanic internal wave signature extraction in the Sulu Sea by a pixel attention U-Net: PAU-Net. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 4000905. http://doi.org/10.1109/LGRS.2022.3230086.
20. Zhang, H.S.; Sun, J.Y.; Qi, K.T.; Zheng, Y.G.; Lu, J.J.; Zhang, Y. Stripe segmentation of oceanic internal waves in SAR images based on SegFormer. *Front. Mar. Sci.* **2025**, *11*, 1456294. http://doi.org/10.3389/fmars.2024.1456294.
21. Duan, B.; Barintag, S.; Meng, J.; Gong, M. Stripe extraction of oceanic internal waves using PCGAN with small-data training. *Remote Sens.* **2024**, *16*, 787. https://doi.org/10.3390/rs16050787.
22. Meng, J.; Sun, L.; Zhang, H.; Hu, B.; Hou, F.; Bao, S. Remote sensing survey and research on internal solitary waves in the South China Sea-Western Pacific-East Indian Ocean (SCS-WPAC-EIND). *Acta Oceanol. Sin.* **2022**, *41*, 154–170. https://doi.org/10.1007/s13131-022-2018-0.
23. Cai, S.; Xie, J. A propagation model for the internal solitary waves in the northern South China Sea. *J. Geophys. Res.-Ocean.* **2010**, *115*, C12074. https://doi.org/10.1029/2010JC006341.
24. Barnes, W.L.; Pagano, T.S.; Salomonson, V.V. Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. *IEEE Trans. Geosci. Remote Sens.* **2002**, *36*, 1088–1100. https://doi.org/10.1109/36.700993.
25. Zhao, Z.; Liu, B.; Li, X. Internal solitary waves in the China seas observed using satellite remote-sensing techniques: A review and perspectives. *Int. J. Remote Sens.* **2014**, *35*, 3926–3946. https://doi.org/10.1080/01431161.2014.916442.
26. Ritter, N.; Ruth, M. The GeoTiff data interchange standard for raster geographic images. *Int. J. Remote Sens.* **1997**, *18*, 1637–1647. http://doi.org/10.1080/014311697218340.
27. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271. http://doi.org/10.1109/CVPR.2017.690.
28. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. http://doi.org/10.1109/CVPR.2016.91.
29. He, L.; Zhou, Y.; Liu, L.; Zhang, Y.; Ma, J. Application of the YOLOv11-seg algorithm for AI-based landslide detection and recognition. *Sci. Rep.* **2025**, *15*, 12421. https://doi.org/10.1038/s41598-025-95959-y.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.
31. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 390–391. https://doi.org/10.1109/CVPRW50498.2020.00203.
32. Zhong, W.; Li, X.; Yue, X.; Feng, W.; Yu, Q.; Chen, J.; Chen, B.; Zhang, L.; Cai, X.; Wen, J. A Classification Method for the Severity of Aloe Anthracnose Based on the Improved YOLOv11-seg. *Agronomy* **2025**, *15*, 1896. https://doi.org/10.3390/agronomy15081896.

33. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 6070–6079. https://doi.org/10.1109/ICCV51070.2023.00558.
34. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
35. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.324.
36. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000. https://doi.org/10.48550/arXiv.1911.08287.
37. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Proceedings of the International Workshop on Deep Learning in Medical Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 240–248. https://doi.org/10.1007/978-3-319-67558-9_28.
38. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520. https://doi.org/10.48550/arXiv.1608.01471.
39. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026.
40. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32. https://doi.org/10.48550/arXiv.1912.01703.
41. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101. https://doi.org/10.48550/arXiv.1711.05101.
42. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Proceedings of the European Conference on Information Retrieval*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359. https://doi.org/10.1007/978-3-540-31865-1_25.
43. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
44. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.
45. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28. https://doi.org/10.48550/arXiv.1506.01497.
46. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475. https://doi.org/10.48550/arXiv.2207.02696.
47. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 12077–12090. https://doi.org/10.48550/arXiv.2105.15203.
48. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306. https://doi.org/10.48550/arXiv.2102.04306.
49. Zhang, X.; Li, X. Constructing a 22-year internal wave dataset for the northern South China Sea: Spatiotemporal analysis using MODIS imagery and deep learning. *Earth Syst. Sci. Data* **2024**, *16*, 5131–5144. https://doi.org/10.5194/essd-16-5131-2024.
50. Li, X.; Zhao, Z.; Pichel, W.G. Internal solitary waves in the northwestern South China Sea inferred from satellite images. *Geophys. Res. Lett.* **2008**, *35*, L13605. https://doi.org/10.1029/2008GL034272.
51. Li, X.; Jackson, C.R.; Pichel, W.G. Internal solitary wave refraction at Dongsha atoll, South China Sea. *Geophys. Res. Lett.* **2013**, *40*, 3128–3132. https://doi.org/10.1002/grl.50614.
52. Huang, X.; Huang, S.; Zhao, W.; Zhang, Z.; Zhou, C.; Tian, J. Temporal variability of internal solitary waves in the northern South China Sea revealed by long-term mooring observations. *Prog. Oceanogr.* **2022**, *201*, 102716. https://doi.org/10.1016/j.pocean.2021.102716.

53. Cai, S.; Xie, J.; Xu, J.; Wang, D.; Chen, Z.; Deng, X.; Long, X. Monthly variation of some parameters about internal solitary waves in the South China Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2014**, *84*, 73–85. http://dx.doi.org/10.1016/j.dsr.2013.10.008.
54. Cai, S.; Xie, J.; He, J. An overview of internal solitary waves in the South China Sea. *Surv. Geophys.* **2012**, *33*, 927–943. https://doi.org/10.1007/s10712-012-9176-0.

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.