Journal of *Marine Science and Engineering*

MDPI

Article

YOLO-SG: Seafloor Topography Unit Recognition and Segmentation Algorithm Based on Lightweight Upsampling Operator and Attention Mechanisms

Yifan Jiang [1,2], Ziyin Wu [1,2,3,4,*], Fanlin Yang [1], Dineng Zhao [1,2,*], Xiaoming Qin [2,3], Mingwei Wang [2] and Qiang Wang [2]

[1] College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China; yfjiang56@163.com (Y.J.); flyang@sdust.edu.cn (F.Y.)
[2] State Key Laboratory of Submarine Geoscience, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China; qinxm@zju.edu.cn (X.Q.); wangmw@sio.org.cn (M.W.); qiangwsio@163.com (Q.W.)
[3] School of Oceanography, Shanghai Jiao Tong University, Shanghai 200240, China
[4] Ocean College, Zhejiang University, Zhoushan 316021, China
* Correspondence: zywu@sio.org.cn (Z.W.); zhaodineng@sio.org.cn (D.Z.)

Abstract: The recognition and segmentation of seafloor topography play a crucial role in marine science research and engineering applications. However, traditional methods for seafloor topography recognition and segmentation face several issues, such as poor capability in analyzing complex terrains and limited generalization ability. To address these challenges, this study introduces the SG-MKD dataset (Submarine Geomorphology Dataset—Seamounts, Sea Knolls, Submarine Depressions) and proposes YOLO-SG (You Only Look Once—Submarine Geomorphology), an algorithm for seafloor topographic unit recognition and segmentation that leverages a lightweight upsampling operator and attention mechanisms. The SG-MKD dataset provides instance segmentation annotations for three types of seafloor topographic units—seamounts, sea knolls, and submarine depressions—across a total of 419 images. YOLO-SG is an optimized version of the YOLOv8l-Segment model, incorporating a convolutional block attention module in the backbone network to enhance feature extraction. Additionally, it integrates a lightweight, general upsampling operator to create a new feature fusion network, thereby improving the model's ability to fuse and represent features. Experimental results demonstrate that YOLO-SG significantly outperforms the original YOLOv8l-Segment, with a 14.7% increase in mean average precision. Furthermore, inference experiments conducted across various research areas highlight the model's strong generalization capability.

Keywords: instance segmentation; seafloor topography and geomorphology; instance segmentation dataset; attention mechanism; generalization ability; YOLOv8

check for updates

Academic Editor: Leszek Chybowski

Received: 23 January 2025
Revised: 1 March 2025
Accepted: 12 March 2025
Published: 16 March 2025

Citation: Jiang, Y.; Wu, Z.; Yang, F.; Zhao, D.; Qin, X.; Wang, M.; Wang, Q. YOLO-SG: Seafloor Topography Unit Recognition and Segmentation Algorithm Based on Lightweight Upsampling Operator and Attention Mechanisms. *J. Mar. Sci. Eng.* **2025**, *13*, 583. https://doi.org/10.3390/jmse13030583

Copyright: © 2025 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The ocean is a vital environment for human activities and sustenance, and the exploration and development of the ocean have always attracted significant attention. The seafloor constitutes a crucial part of the ocean [1]. As the foundation of the ocean, the seafloor is a core area for key processes such as global plate tectonics, ocean current movement, and sediment deposition. The seafloor environment records important geological history, such as the trajectory of crustal movements, long-term trends in climate change, and the evolution of deep-sea ecosystems. In addition, the seafloor harbors a vast amount

of untapped resources that have significant economic and strategic value [2]. Studying the seafloor is not only key to understanding the overall ocean system but also an essential requirement for promoting resource development, disaster early warning, and environmental protection [3]. In seafloor research, the study of seafloor topography and geomorphology is particularly important. Seafloor topography and geomorphology directly influence the distribution of seafloor ecosystems, ocean current paths, and sediment movement patterns. Classifying geomorphological units such as mid-ocean ridges, ocean trenches, deep-sea plains, and seamounts can reveal the seafloor's structure and resource distribution characteristics.

Before the emergence of modern surveying technologies, the study of seafloor topography and geomorphology relied on manual detection and hand-drawn mapping. Scholars classified seafloor topography and geomorphology based on limited acoustic survey data and submersible observations. These methods were unsystematic, characterized by scarce and low-precision data, and primarily suited for describing localized topographic features. Classical classification methods are based on expert experience, relying on visual inspection and known geomorphological features [4]. Advancements in science and technology have led to the emergence of modern surveying techniques. Multi-beam echo sounders and side-scan sonars [5–8] were applied to underwater detection, and scholars started researching rule-based segmentation methods. At this time, classification mainly relied on simple topographic features such as depth, elevation, slope, and curvature. Different geomorphological units could be categorized by setting thresholds. Harris et al. [9] proposed a global seafloor geomorphology classification framework using slope, elevation, and curvature as the main classification features. Luiceer et al. [10] used the object-based image analysis method, combining multi-beam data and echo intensity data, to perform automated segmentation by setting thresholds and rules. While these methods are straightforward, they lack the capability to analyze complex terrains. Automated segmentation methods based on statistical analysis have also been introduced for seafloor topography classification [11] to improve accuracy. Such methods achieve automatic classification of geomorphology by analyzing the data distribution patterns of multi-dimensional topographic features. Common methods include cluster analysis (such as K-means clustering), principal component analysis, and Gaussian mixture models. Jung et al. [12] proposed a clustering method based on inverse histograms that detects peaks and optimizes segmentation parameters to achieve segmentation of topographic point clouds. Cutter et al. [13] successfully divided the seafloor into multiple classes by extracting texture features from different regions and using fuzzy K-means clustering. These methods improve the accuracy and automation of geomorphological classification via statistical analyses of large datasets but still have limitations when dealing with complex terrains. With the improvement in computational power, machine-learning methods such as support vector machines, decision trees, and random forests have been introduced into seafloor topography and geomorphology classifications [14–18]. Machine-learning methods can automatically recognize topographic features by training models, improving classification accuracy. Researchers can label topographic data based on multi-dimensional feature data and then train models to learn the relationships between features and geomorphology, enabling automated classification. Masetti et al. [19] combined the use of deep-sea detection digital elevation models (DEMs) and acoustic backscatter mosaics to segment the seafloor. Although machine-learning methods have demonstrated potential and application value in seafloor topography and geomorphology, they have drawbacks, such as a dependence on high-quality data and a limited generalization ability.

In recent years, with its rapid development and widespread application, deep learning has shown tremendous advantages in image recognition and segmentation, especially

convolutional neural networks, which have performed exceptionally well in processing imagery and point cloud data [20–23]. Deep-learning models can undergo end-to-end training with large datasets, automatically learning complex features without relying on manual feature engineering. They have the advantages of good adaptability to complex nonlinear data, a strong generalization ability, and compatibility with multi-source data. However, current research on the automatic recognition and segmentation of seafloor topography and geomorphology using deep-learning methods is relatively limited, with most studies focusing on sediment classification. Therefore, the application of deep learning to seafloor topography and geomorphology segmentation is still immature, facing several prominent issues, such as the lack of mature models, the scarcity of available datasets, and the segmentation scale being too large to effectively apply to secondary or smaller geomorphological units.

YOLO (you only look once: unified, real-time object detection) is an object detection system based on a single neural network proposed by Joseph Redmon, Ali Farhadi, and others in 2015. In 2017, at CVPR, Joseph Redmon and Ali Farhadi published YOLOv2, which further improved detection accuracy and speed [24]. YOLOv8 builds upon historical versions of the YOLO series and introduces new features and improvements to further enhance its performance and flexibility, making it the best choice for tasks such as object detection, image segmentation, and pose estimation. Its simplicity, speed, and efficiency make it one of the most popular algorithms in computer vision. YOLO-Segment is an extension of the YOLO detection framework designed to handle instance segmentation tasks [25]. Standard YOLO focuses on detecting and classifying objects using bounding boxes, while YOLO-Segment provides pixel-level masks for individual objects.

To address the challenges faced by deep-learning methods in seafloor topography and geomorphology recognition and segmentation, this paper uses YOLOv8l-Segment as the baseline network [26] and proposes a seafloor topography and geomorphology image segmentation algorithm based on lightweight upsampling and attention mechanisms (YOLO-SG). Additionally, a dataset called SG-MKD is created. This method optimizes and improves the YOLOv8l-Segment network, enabling automatic recognition and segmentation of geomorphological units, such as seamounts, sea knolls, and submarine depressions. Compared with the baseline network YOLOv8l-Segment, YOLO-SG achieves a 14.7% improvement in the mean average precision, demonstrating superior detection performance and a stronger ability to capture small targets.

2. Methods

To achieve automatic segmentation of seafloor geomorphological units, this study improves on the YOLOv8l-Segment, resulting in a network, YOLO-SG, that performs better in the topography and geomorphology segmentation task. As shown in Figure 1, the original YOLOv8l-Segment network consists of three parts: the backbone (main network); the head (output layer); and the neck (intermediate layer). The backbone, the foundation of the model, is responsible for extracting features from the input image. These features serve as the foundation for the subsequent network layers to perform object detection and segmentation. The neck, responsible for multi-scale feature fusion, enhances feature representation by combining feature maps from different stages of the backbone. The head, responsible for the final object detection and classification tasks, includes a detection head and a classification head. YOLO-SG introduces the convolutional block attention module (CBAM) in the backbone network, enhancing the model's feature extraction capabilities. CBAM improves model performance via adaptive feature recalibration and a dual attention mechanism that includes both channel attention and spatial attention. YOLO-SG introduces the CARAFE operator in the neck network, improving the structure and forming a new

network architecture. The CARAFE operator has characteristics such as a large receptive field, sampling kernels that are semantically related to the feature maps, and a lightweight design. It enhances the feature fusion effect of the network without introducing excessive computational overhead. Figure 2 shows the network structure of YOLO-SG.

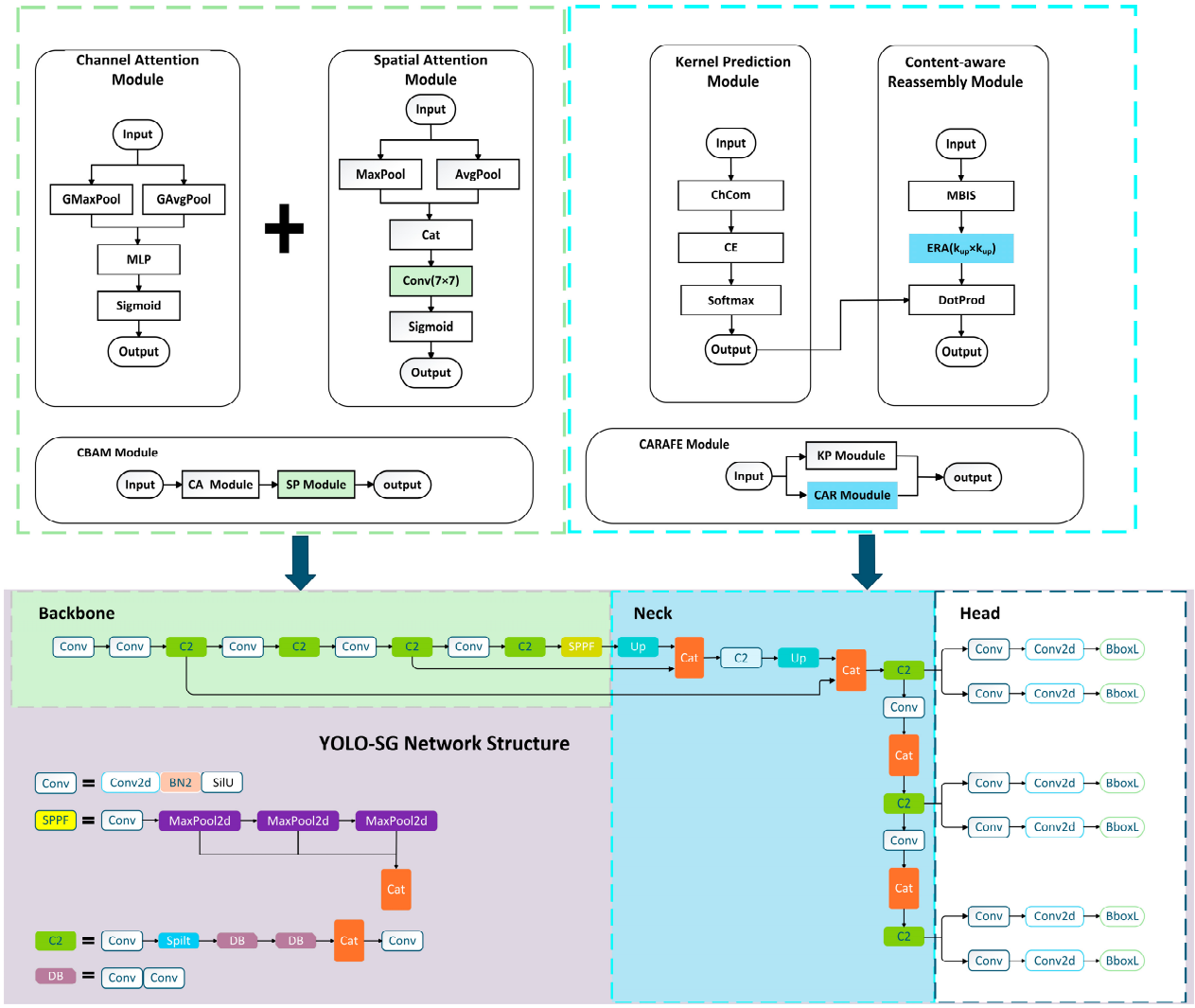


Figure 1. Original architecture diagram of YOLOv8l-Segment. The convolutional block attention module (CBAM) is introduced in the backbone network, while CARAFE is integrated into the neck network, forming a new network architecture.

2.1. Feature Extraction Module Integrated with CBAM

YOLOv8l-Segment was selected as the baseline model for this experiment to achieve the most accurate identification and segmentation of the topography and geomorphology. This model has a significant advantage in accuracy; however, because of its large size, it also consumes considerable computational resources compared with lightweight models. Given the limited number of seafloor topography and geomorphology image samples and their sparse features, the model needs to fully exploit the valuable information within these limited samples to improve the recognition performance as much as possible. Therefore, improvements to the model need to consider two aspects simultaneously: (1) maximizing the feature extraction capability of the backbone network and (2) minimizing computational consumption while achieving a certain level of performance improvement to achieve a high computational efficiency ratio. Accordingly, this study introduces the CBAM attention mechanism into the backbone network and proposes a new backbone network structure.

This backbone network structure can effectively enhance the network's feature extraction capability with a minimal increase in computational consumption.

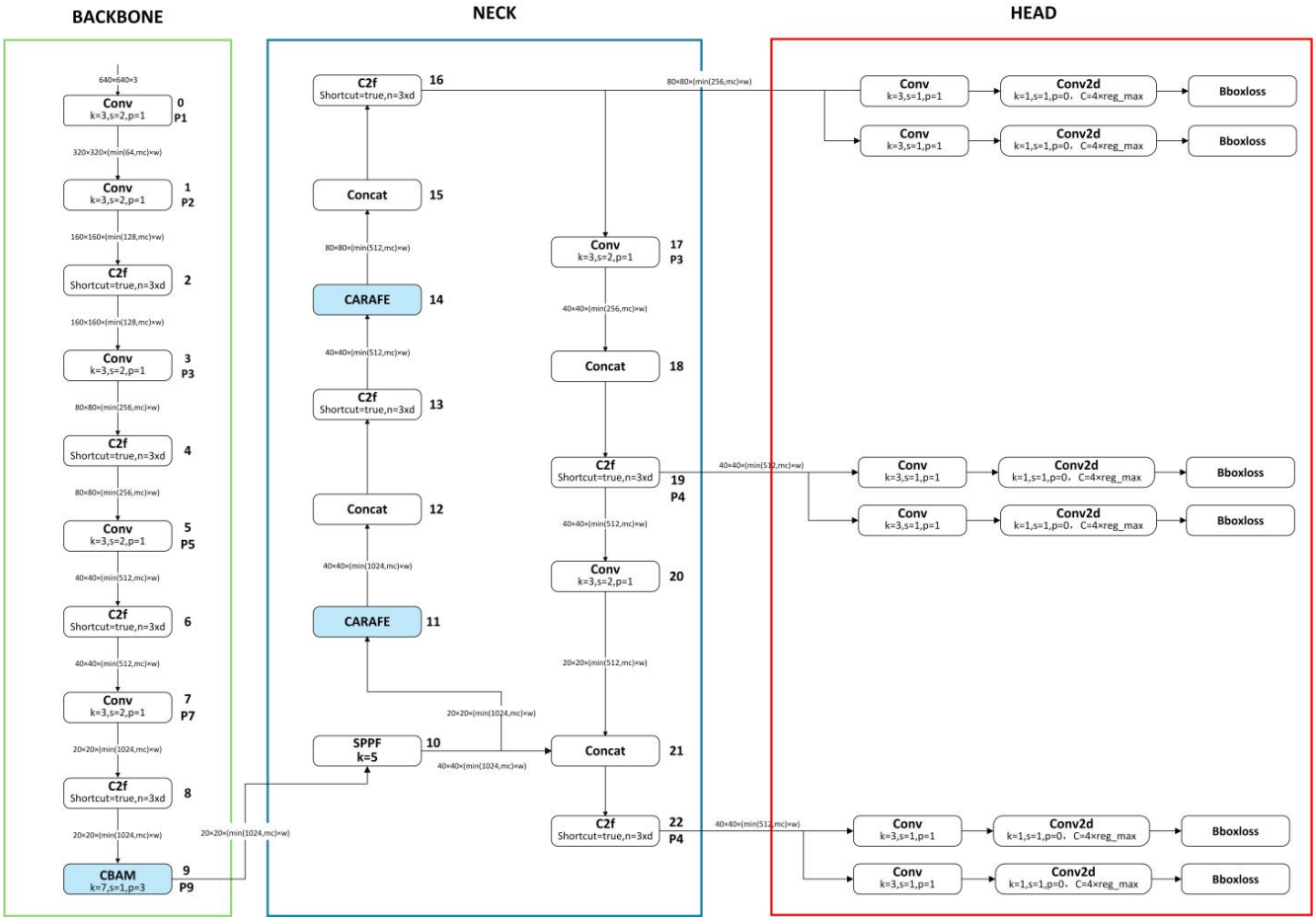


Figure 2. Improved network structure of YOLO-SG.

CBAM was proposed by Woo et al. [27] in 2018. The design of CBAM introduces an attention mechanism that allows the network to focus more effectively on important channel and spatial location information, thereby enhancing performance in tasks such as object detection, image recognition, and semantic segmentation.

CBAM consists of two main sub-modules, the channel attention module (CAM) [28] and the spatial attention module [29], that sequentially enhance the input features. The overall structure of CBAM is shown in Figure 3. The input feature $F \in R^{C\times H\times W}$ represents a tensor with C channels, H height, and W width, where each element belongs to the real number domain R. This input first passes through the Channel Attention Module (CAM), where a one-dimensional convolution kernel $M_c \in R^{C\times 1\times 1}$ is applied. The convolution result of M_C is multiplied by the original input F, and the resulting feature is then passed as input to the spatial attention module. In the spatial attention module, a two-dimensional convolution $M_s \in R^{C\times H\times W}$ is applied, and the output of this operation is also multiplied by the original input F.

$$\begin{cases} F' = MC(F) \otimes F \\ F'' = MC(F') \otimes F' \end{cases} \quad (1)$$

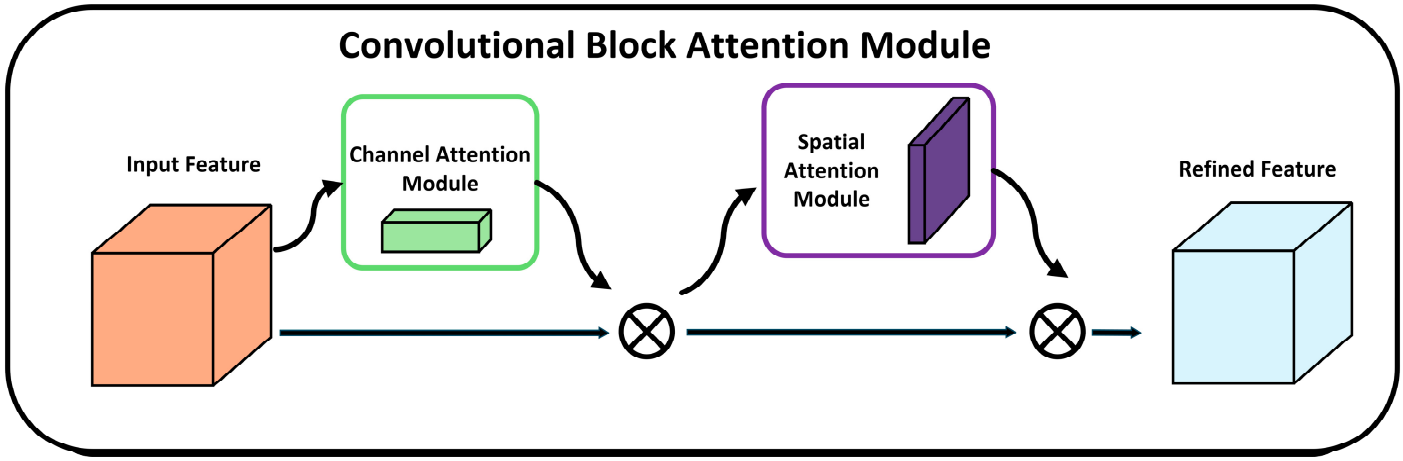


Figure 3. Schematic diagram of the CBAM structure.

(1) Channel attention module (CAM)

In CAM, the channel dimension remains unchanged, while the spatial dimension is compressed. This module focuses on the meaningful information in the image.

As shown in Figure 4, the input feature map is passed through two parallel MaxPool and AvgPool layers, reducing the spatial dimensions from C × H × W to C × 1 × 1. Then, it goes through the shared multi-layer perception module. In this module, the channel number is first compressed to $1/r$ (where r is the reduction factor) of the original value and then expanded back to the original channel number. The outputs are activated using the rectified linear unit (ReLU) activation function, resulting in two activated outputs. These two outputs are then element-wise added together, followed by a sigmoid activation function to obtain the channel attention output. Finally, this output is multiplied by the original feature map to restore the dimensions back to C × H × W [28].

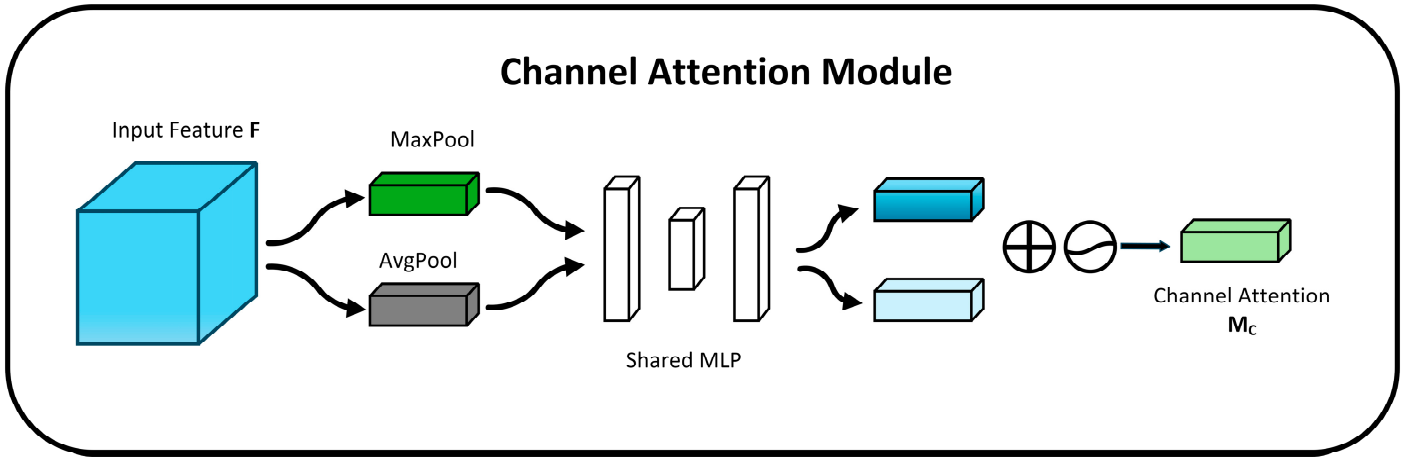


Figure 4. Schematic diagram of the structure of the channel attention mechanism.

Channel attention mechanism calculation formula:

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1(W_0(F^c_{max}))\right). \end{aligned} \tag{2}$$

(2) Spatial attention module

In the spatial attention module, the spatial dimension remains unchanged while the channel dimension is compressed. This module focuses on the positional information of

the object in the image. As shown in Figure 5, the output of the channel attention is passed through MaxPooling and AveragePooling to obtain two feature maps of size 1 × H × W. Then, the two feature maps are concatenated using the concat operation, and finally, the output is multiplied by the original feature map to restore its dimensions back to C × H × W [29].

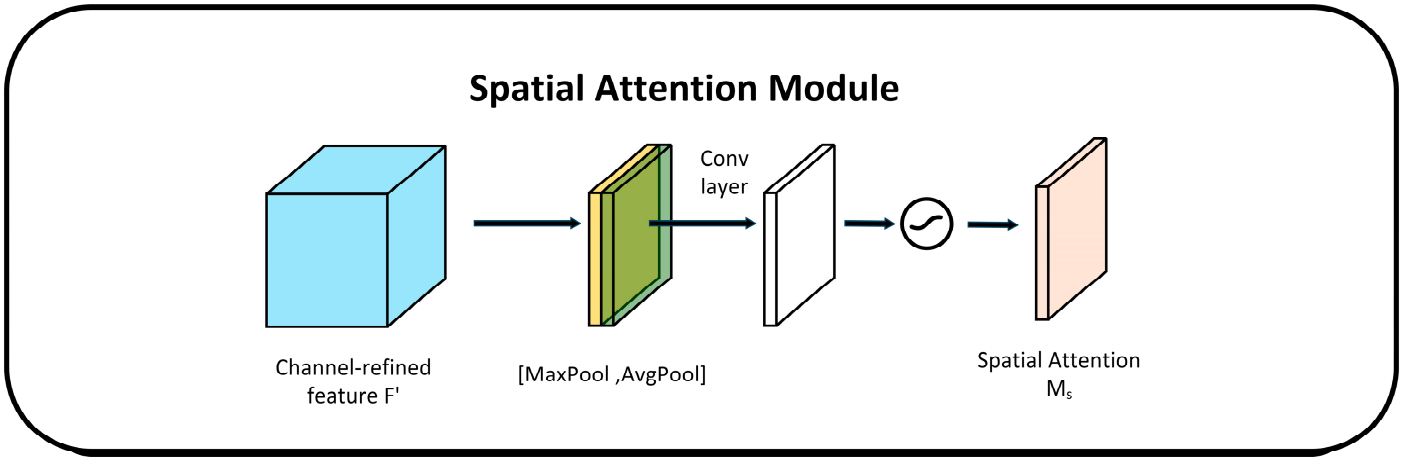


Figure 5. Schematic diagram of the structure of the spatial attention mechanism.

Spatial attention mechanism calculation formula:

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) = \sigma\left(f^{7\times7}\left[F^s_{avg}; F^s_{max}\right]\right). \quad (3)$$

2.2. Feature Fusion Network Integrated with the CARAFE Operator

The contours of seabed geomorphic units are complex. Although incorporating CBAM in the backbone network improves the model's feature extraction capability for seabed topographic images to some extent, the simplicity of the network structure may prevent it from fully leveraging the feature information in seabed topographic images. Additionally, seabed geomorphic units are characterized by their complex contours. During the process of converting DEMs into seabed topographic maps, significant scale differences between various seabed geomorphic units inevitably result in the presence of small targets. Accordingly, this study introduces the CARAFE module to construct a new feature fusion network, enhancing YOLO-SG's ability to segment small targets and complex shapes. CARAFE generates dynamic reassembly kernels for each pixel via learning rather than using fixed upsampling operations, as illustrated in the kernel prediction module in Figure 6. For small and complex targets, this dynamic kernel generation approach can capture finer features, avoiding issues in traditional methods where small targets are easily overlooked, or the edges of complex shapes appear blurred [30]. In addition, CARAFE performs feature-weighted summation within a local region, effectively aggregating local information, as illustrated in the content-aware reassembly module in Figure 6. This approach enables better preservation of boundary features for complex shapes while assigning higher weights to critical feature points of small targets, thereby improving the segmentation performance. Especially in complex backgrounds, information concerning small targets is easily overshadowed by background features. However, CARAFE enhances the feature contrast of small targets via content-aware reassembly.

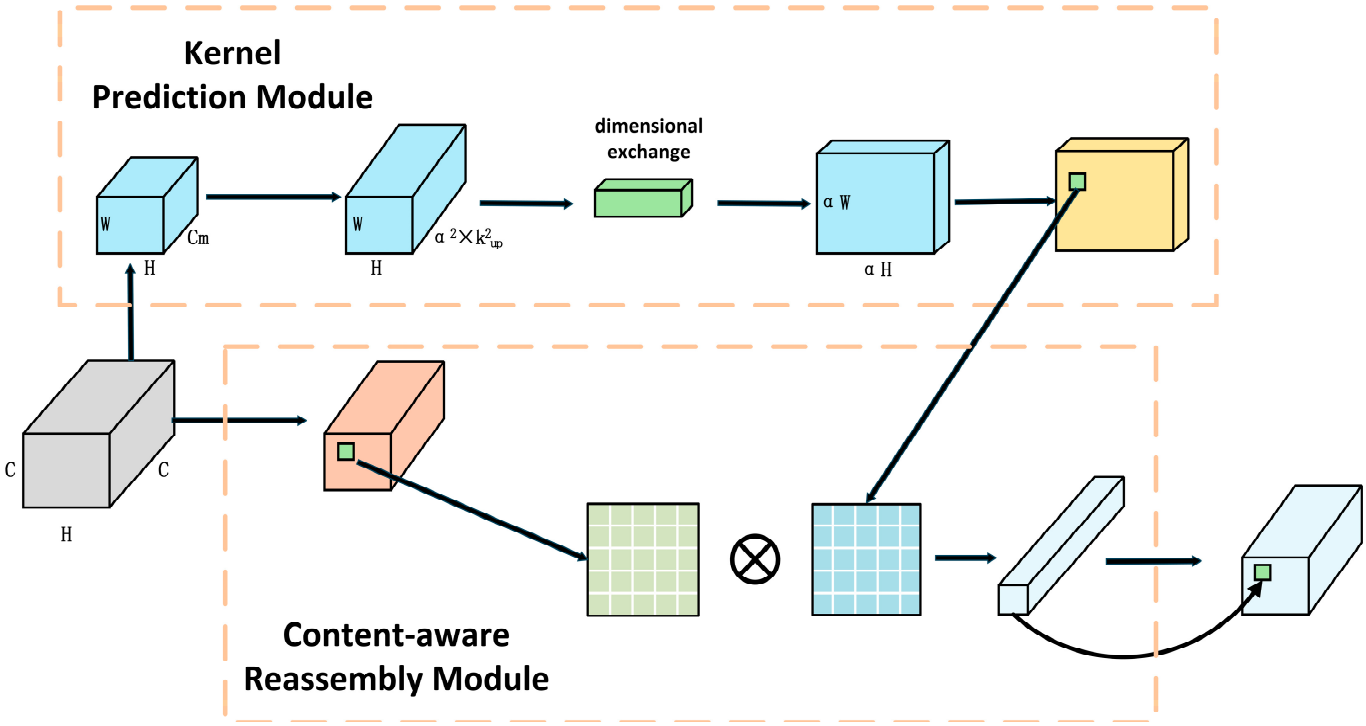


Figure 6. Schematic diagram of the CARAFE operator structure.

3. Dataset

Most current studies on submarine geomorphology recognition and segmentation are rule-based, with limited research in deep learning and a lack of available datasets. For this study, the authors created a dataset called SG-MKD.

Existing terrain models can be categorized into DEM (Digital Elevation Model), DSM (Digital Surface Model), DTM (Digital Terrain Model), TIN (Triangulated Irregular Network), LiDAR (Light Detection and Ranging) point cloud data, and VTM (Vector Terrain Model). In seabed terrain research and applications, DEM is the most widely used model [31–34]. This model records seabed elevation values in a regular grid format, allowing for precise representations of continuous seabed terrain features, such as trenches, seamounts, abyssal plains, and rift valleys. It has many advantages, such as being easy to analyze and process, supporting multi-source data, having abundant data resources, and offering a wide range of application scenarios [35]. The seafloor terrain grids used to generate the DEM in this study were Topo_27.1 and GEBCO 2023 Grid. The Topo model was developed by Professor Sandwell's team at the Scripps Institution of Oceanography (SIO) and has been continuously updated since the 1990s, with a new version released annually. The model has been continuously optimized via improved editing, updated gravity models, and the addition of data, with the Topo_27.1 version released in 2023 [36,37]. This version has an average resolution of 1′ and a latitude range of ±80.738°, allowing clear visibility of seamounts with a size of 1–2 km. The General Bathymetric Chart of the Oceans (GEBCO) is an ocean depth mapping project jointly led by the International Hydrographic Organization and the Intergovernmental Oceanographic Commission. GEBCO_2023 was released in April 2023 with a spatial resolution of 15″. It covers the region from 89°59′52.5″ N to 89°59′52.5″ S and from 179°59′52.5″ W to 179°59′52.5″ E. The model is based on SRTM15+ v2.5.5, integrating seafloor topography information inverted from the SIO Gravity Model V32 and gridded bathymetric datasets developed by the four Seabed2023 regional centers. Both of the above-mentioned topographic grids can be downloaded from their respective official websites.

Seamounts are isolated mountains or peaks on the ocean floor, typically formed by volcanic activity, with significant protrusions beneath the sea surface. The base diameter of seamounts typically ranges from 10 km to 100 km, with some large seamounts having diameters that can reach several hundreds of kilometers [38,39]. Sea knolls are relatively smaller underwater mountains, typically resulting from the degradation of larger seamounts or small-scale tectonic features. They are often found in continental slopes or abyssal plain areas. The base diameters of these seamounts are usually several kilometers to tens of kilometers, smaller than typical seamounts. In this study, the grid is divided into dimensions of $2^\circ \times 2^\circ$, which ensures that a single image can fully accommodate large geomorphic units while avoiding the blurriness of smaller geomorphic units, which could affect recognition and segmentation.

LabelMe is an open-source tool created by the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology in 2008, specifically for image annotation and semantic segmentation. It provides a platform that allows users to manually annotate images and create label data for training machine-learning models or conducting computer vision research [40]. The SG-MKD dataset was annotated using LabelMe for the labeling.

The dataset contains 419 images with recognized and segmented objects, including three types of seabed features: seamounts; sea knolls; and submarine depressions. Each image contains at least one type of seabed feature, and the image resolution is uniform at 120×120 pixels. Several sample images from the dataset are shown in Figure 7.

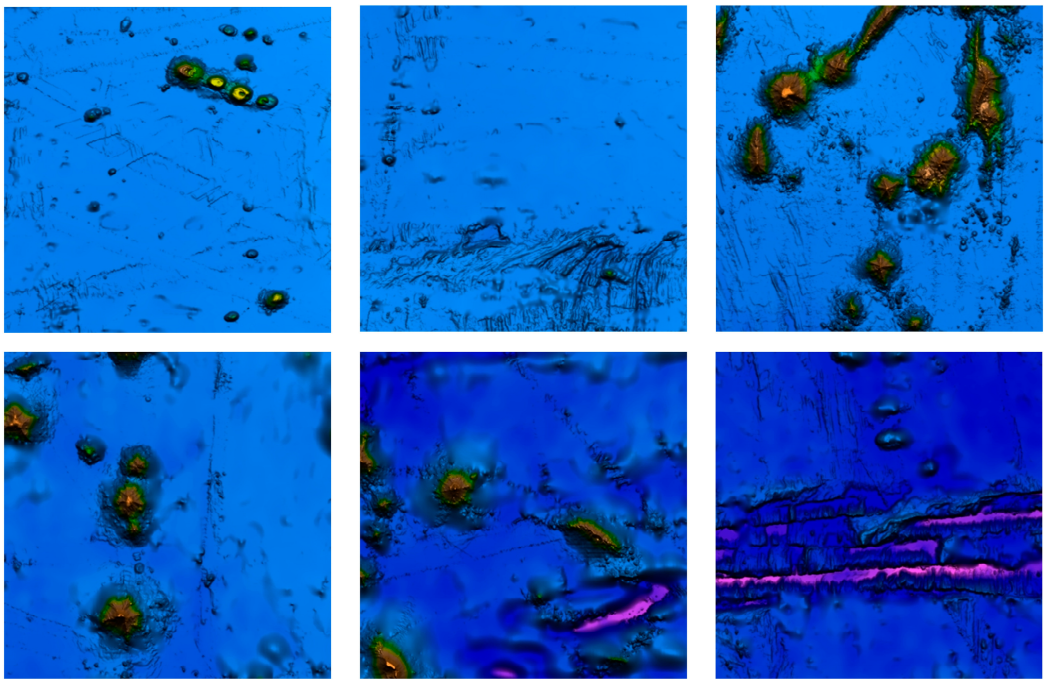


Figure 7. Display of partial dataset images (The image enlarged from the original 120×120 size).

Instance segmentation of seabed geomorphic units is a challenging task. During the annotation process, it is necessary to rely on expert knowledge to determine the type of seabed feature and, at the same time, accurately mark the boundaries of each seabed geomorphology type. This makes the creation of a dataset extremely labor-intensive. Consequently, this study performed data augmentation on the SG-MKD dataset. Data augmentation [41] plays a crucial role in improving the performance and robustness of a model. Introducing various variants into a dataset helps the model better understand

the complex task of image detection and instance segmentation. The key roles of data augmentation in instance segmentation are as follows.

1. Enhancing Generalization: Introducing variability through augmentation helps prevent the model from overfitting to specific features of the training set. It improves the model's ability to segment objects under different conditions, increasing the feature richness of the dataset [42] and ensuring that the model performs well on unseen data.
2. Increasing Dataset Size [43]: Instance segmentation tasks require pixel-level annotations, which are time-consuming and expensive to generate. Augmentation can effectively expand the training data without the need for additional labeling work.
3. Improving Mask Prediction Accuracy: Augmented examples with different shapes, rotations, or scales can help the model learn the accurate mask boundaries of objects under different orientations or distortions [44].
4. Improved Detection of Small and Occluded Objects [45]: Techniques such as enlarging smaller objects or introducing partial occlusion during training can help the model address challenging problems, such as detecting small objects in real-world scenarios.

To increase the target quantity and characteristics of the SG-MKD dataset, data augmentation was performed using techniques such as adding noise, adjusting brightness, randomly changing pixel values, translating, and flipping. The total number of augmented images in the dataset is 1670. The dataset was divided into training sets and validation sets according to a ratio of 8:2. Note that the dataset partitioning was performed prior to data augmentation.

4. Experimental Results

All experiments in this study were conducted on a computer equipped with an Intel(R) Core(TM) i9-10900X CPU, NVIDIA GeForce RTX 3090 GPU, and 64 GB RAM.

The experiments were conducted on a Windows 10 64-bit system using the PyTorch framework, with Python version 3.8, PyTorch version 1.12.0, and CUDA version 11.6. The model training parameters were set as follows: the initial learning rate was set to 0.01; the weight decay was set to 0.0005; the momentum was set to 0.937; the batch size was set to 16; the image size was set to 120 × 120; and the total number of epochs for training was 200.

4.1. Evaluation Indicators

In instance segmentation, evaluation metrics are designed to assess the model's ability to detect objects and accurately predict their segmentation masks. These metrics combine aspects of both object detection and semantic segmentation evaluation. The detection is considered a positive sample by calculating the intersection over union (IoU) between the predicted mask and the ground truth mask in the image. The corresponding detection results are defined as follows: true positive (*TP*)—the number of correctly detected positive samples; false positive (*FP*)—the number of negative samples incorrectly detected as positive; and false negative (*FN*)—the number of positive samples incorrectly detected as negative. In this instance segmentation task, which involves three categories, the IoU threshold was set to 0.5. To objectively evaluate the accuracy of the model, the average precision (*AP*) for each category and the mean average precision (*mAP*) across all categories were used as evaluation metrics. Their formulas are as follows:

$$AP = \int_0^1 P(R)dR, \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i. \tag{5}$$

In this context, precision (P) refers to the probability of correctly detecting positive samples among all detected positive samples, while recall (R) represents the probability of correctly identifying positive samples among all actual positive samples. Their formulas are as follows:

$$P = TP/(TP + FP), \tag{6}$$

$$R = TP/(TP + FN). \tag{7}$$

To assess the computational complexity of the model, this paper uses the floating-point operations (FLOPs), number of parameters (Params), model file size (Size), and detection speed (Speed) as evaluation metrics.

4.2. Results

4.2.1. Improvement Experiments

To demonstrate the effectiveness of the proposed method, improvement experiments were conducted on the basis of the research method presented in Section 2 using the SG-MKD dataset, the experimental results of which are presented here. First, the YOLOv8l-Segment was used as the baseline model to train and validate the experimental dataset. Then, the proposed YOLO-SG was experimentally validated. The detection and segmentation results of the two networks were compared, and the experimental results are shown in Table 1.

Table 1. Computational complexity and detection accuracy of YOLOv8l-Segment and YOLO-SG on the SG-MKD dataset (bold text is meant to highlight the enhancement of the model's performance).

Model	Class	Precision/%	Recall/%	AP/%	mAP/%	Params/M	FLOPs/G	Size/MB
YOLOv8l-Segment	the mound	64.8	73.5	74.0	73.7	45.93	220.8	92.3
	mountain	77.0	88.6	82.7				
	Submarine depressions	66.5	58.1	64.4				
YOLO-SG	the mound	84.8	83.5	**88.9 ↑14.9**	**88.4 ↑14.7**	**46.90 ↑0.97**	**221.8 ↑1.8**	**94.3 ↑2.0**
	mountain	97.0	98.6	**97.3 ↑14.6**				
	Submarine depressions	86.5	66.1	**82.8 ↑18.4**				

In Table 1, it can be observed that YOLOv8l-Segment exhibits a significant imbalance in detection accuracy across the three categories. The accuracy for seamounts is relatively high; however, the detection accuracies for sea knolls and submarine depressions are lower, with submarine depressions having an average precision of only 64.4%. This phenomenon is primarily attributed to the variation in the target scales and the relatively small number of original samples for these two categories, which significantly impacted the target detection accuracy.

From Table 1, it is evident that compared with the YOLOv8l-Segment, the proposed YOLO-SG demonstrates a significant improvement in detection accuracy. Specifically, the *mAP* value increased by 14.7%, and the *AP* values for seamounts, sea knolls, and seabed depressions improved by 14.9%, 14.6%, and 18.4%, respectively. At the same time, YOLO-SG's parameter count increased by only 0.97 M, and its FLOPs grew by 1.8 G. This demonstrates that the model achieved a significant improvement in accuracy while incurring only a slight increase in computational costs. Figure 8 shows the original images and the detection results of YOLOv8l-Segment and YOLO-SG on two seafloor topographic images. Compared with YOLOv8l-Segment, YOLO-SG demonstrates more accurate boundary delineation and effectively reduces the occurrence of missed detections in the same side-scan sonar images. In the feature map comparison shown in Figure 9, YOLO-SG demonstrates better attention to both local and global features during feature

extraction, with a stronger capability for capturing small targets. In terms of feature fusion and representation, YOLO-SG better focuses its attention on the regions of positive samples while reducing attention to irrelevant information such as blurred backgrounds and shadows.

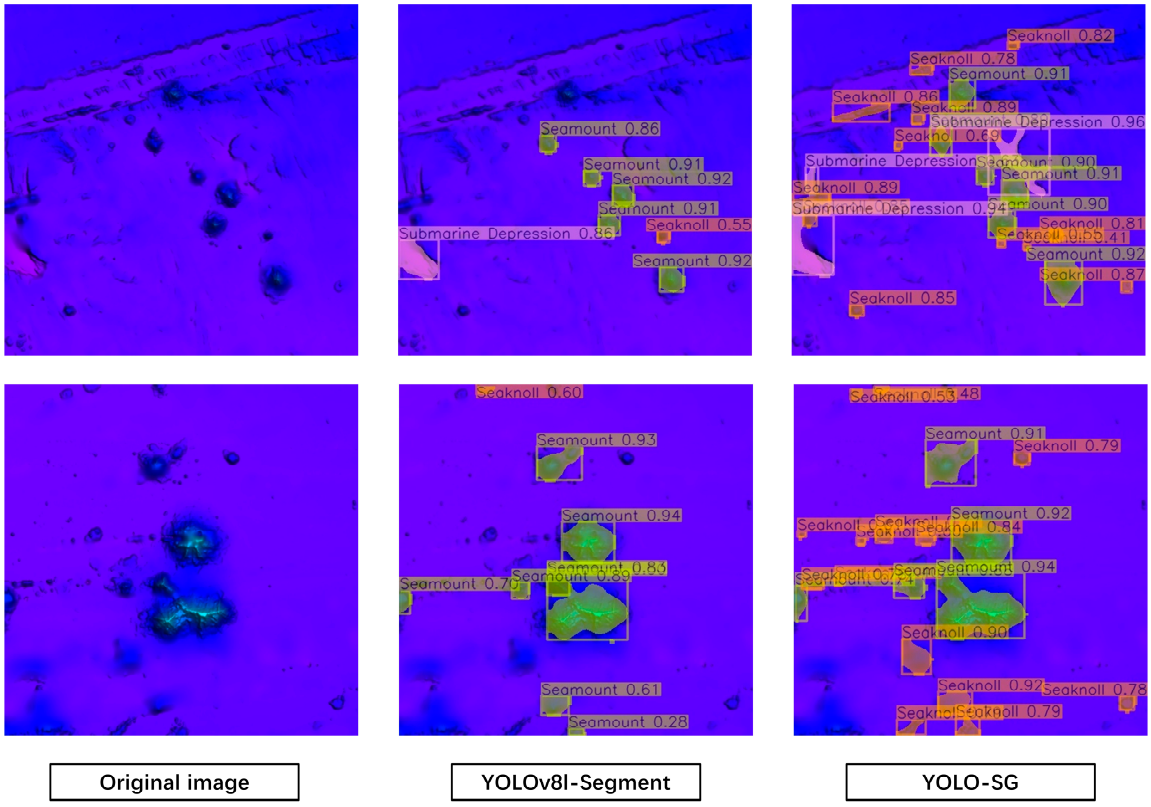


Figure 8. Comparison of seafloor topographic images and the detection results of YOLOv8l-Segment and YOLO-SG on seafloor topographic images.

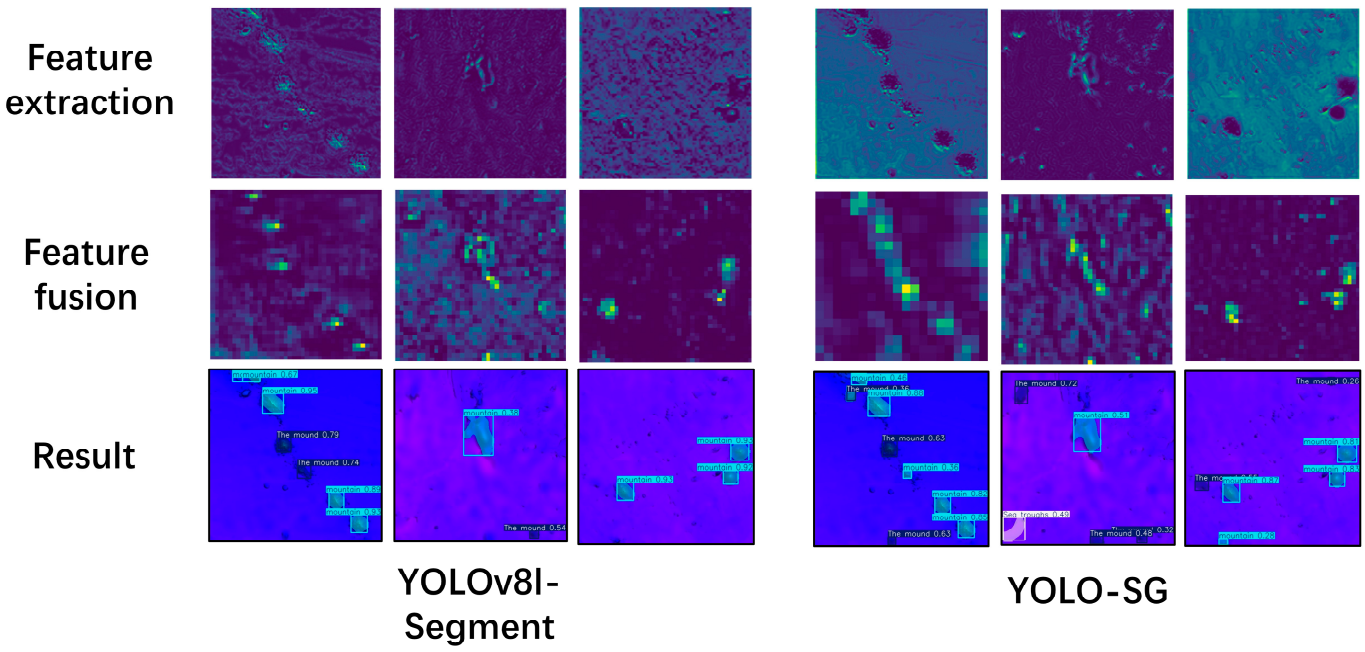


Figure 9. Comparison of feature maps between YOLOv8l-Segment and YOLO-SG for partial images.

4.2.2. Generalization Experiments

Section 3 outlined the numerous challenges encountered during the creation of instance segmentation datasets. As a seafloor geomorphological unit instance segmentation dataset, SG-MKD is similarly constrained by these issues and exhibits certain limitations. The main issues include the uneven distribution of the dataset, primarily concentrated in the North Atlantic, and an insufficient number of samples. These issues limit the reliability of experimental results in other regions. In addition, the fixed input size of 120 × 120 for the training set raises concerns about the algorithm's segmentation and detection performance when handling inputs of different sizes. To demonstrate the good generalization capability of YOLO-SG and its ability to perform seabed geomorphological unit segmentation and detection tasks with different input sizes, we conducted experiments on five randomly selected regions worldwide based on the research methodology proposed in Section 2; the experimental results are presented here.

Research area A is located in the central South Pacific region and is characterized by a dense distribution of seamounts and sea knolls with significant topographic variation. The water depth ranges between 2500 m and 4500 m, and the input size is 1354 × 1283. Research area B is located near the equator in the Mid-Atlantic Ridge region, exhibiting distinct features of linear rift valleys and spreading centers. The water depth ranges between 2500 m and 4500 m, and the input size is 935 × 793. Research area C is located near the equatorial region of Africa, adjacent to the deep-sea area off the west coast of Africa. This region features relatively moderate topographical variations, greater water depths, and isolated underwater mountains and hills. The input size is 906 × 769. Research area D is located in the central Indian Ocean region and is characterized by complex topography, including underwater mountain ranges, hills, and rift valleys. The input size is 1454 × 1171. Research area E is located in the Pacific waters east of Australia, featuring deep-sea basins and underwater mountains. The input size is 1902 × 1768. Figure 10 shows a distribution map of the research areas, and Figure 11 shows the segmentation results of the research areas.

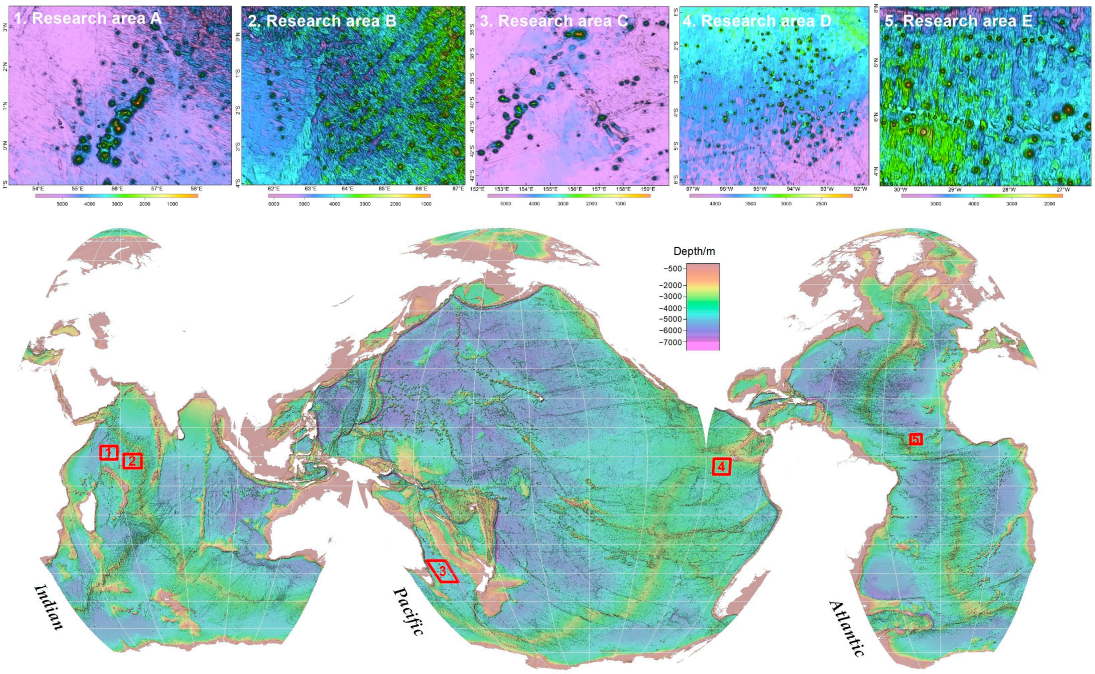


Figure 10. Distribution map of the research areas.

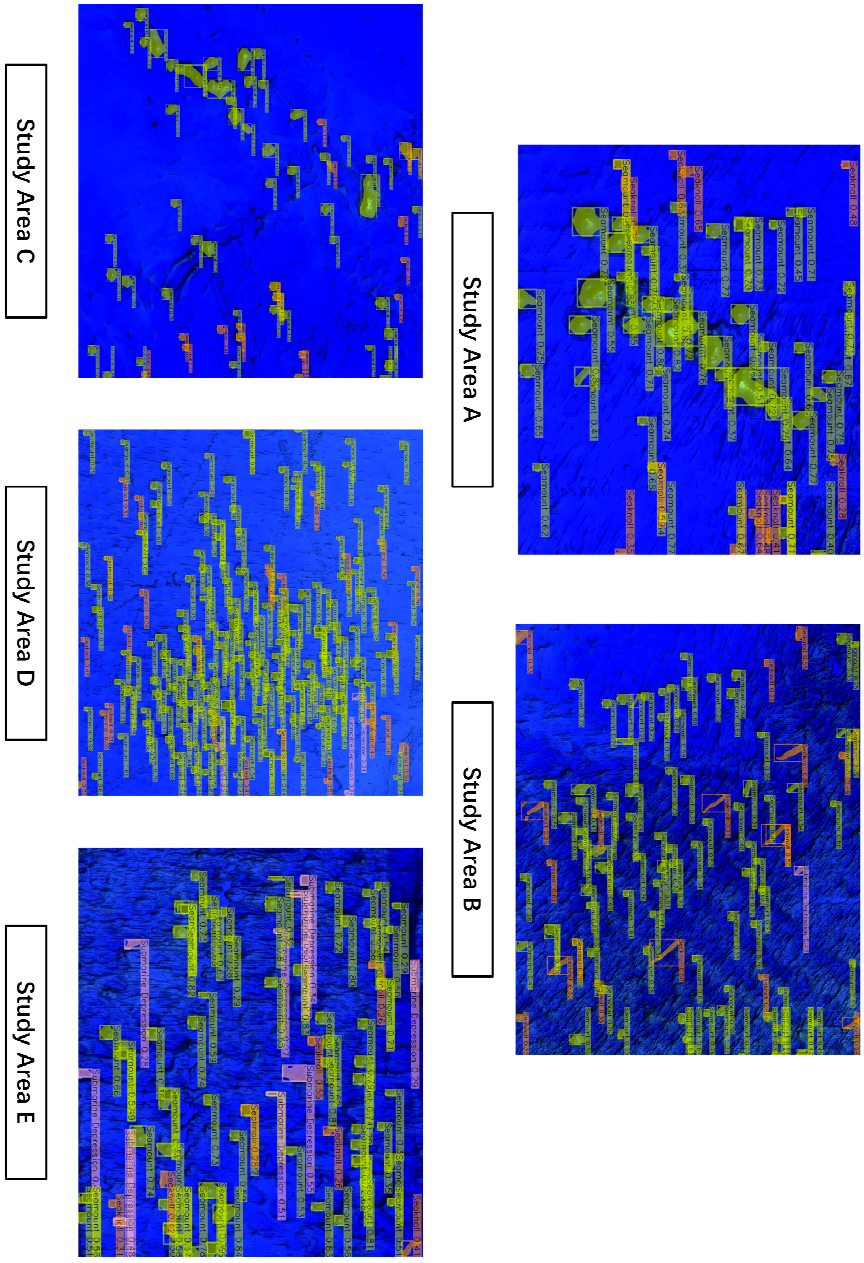


Figure 11. Segmentation results of the research areas.

As shown in Figure 10, the research areas are relatively evenly distributed across the world's oceans, and the surface area of each research area varies. In area A, the basic terrain is relatively flat, with a large number of seamounts and sea knolls, and the seabed geomorphic units are relatively independent. The basic terrain in area B is relatively rugged, with seamounts and guyots as the main geomorphic units, and each unit is relatively independent. The basic terrain in area C is relatively flat; however, the distribution of the geomorphic units is more concentrated. The lower-left part of the basic terrain in area D is flat, while the rest of the area is more rugged, containing numerous geomorphic units. The basic terrain in area E is flat, with the geomorphic units concentrated in the lower left. These five research areas, therefore, cover a variety of terrain features.

As shown in Figure 11, YOLO-SG shows excellent performance in detecting and segmenting seabed geomorphological units under various terrain conditions, which fully demonstrates its strong generalization ability.

4.2.3. Ablation Experiments

To visually demonstrate the impact of each improvement made by the YOLO-SG algorithm on its performance, we conducted ablation experiments on the SG-MKD dataset. The experiment used YOLOv8l-Segment as the baseline model, sequentially adding the CBAM attention mechanism and the CARAFE lightweight upsampling operator as improvement modules. The results are shown in Table 2. The addition of CBAM in Method 1 improved *mAP* by 8.1%, with only a slight increase in the number of parameters and FLOPs. Method 2, which introduces CARAFE to the baseline network, improved *mAP* by 9.3% and increased the parameters by 0.4 M. Finally, YOLO-SG, which incorporates both CBAM and CARAFE, achieved a 14.7% increase in *mAP* compared with the baseline network, with an increase of 1.0 M in the parameters.

Table 2. Changes in the model metrics in the ablation experiment.

Model	Params/M	FLOPs/G	Size/MB	Speed/ms	Precision/%	Recall/%	mAP/%
YOLOv8l-Segment	45.93	220.8	92.3	16.0	75.2	80.9	73.7
Method 1	46.72 ↑0.79	221.4	93.9	16.6	83.9	80.5	81.8 ↑8.1
Method 2	46.11 ↓0.61	221.2	92.7	16.2	87.2	78.5	83.0 ↑2.2
YOLO-SG	46.90	221.8	94.3	16.8	90.9	84.6	88.4 ↑5.4

4.2.4. Comparative Experiments

To better demonstrate the improvements of the proposed instance segmentation method, this section compares it with several State-of-the-Art instance segmentation models on the SG-MKD dataset. The algorithms for comparison include YOLOv5 [46], ASF-YOLO [47] (an improvement based on YOLOv5), Mask R-CNN [46], and Cascade Mask R-CNN [48].

Table 3 presents a comparison of YOLO-SG with other algorithms in terms of the number of parameters, FLOPs, accuracy, and model size. As shown in the table, ASF-YOLO is more lightweight compared to YOLOv5l-Segment, with a notable improvement in accuracy, although it is still not ideal. Cascade Mask R-CNN shows a significant improvement in accuracy compared to Mask R-CNN, achieving a mAP of 86.4%. However, it suffers from high computational complexity. In comparison, YOLO-SG exhibits a significantly lower computational complexity while achieving a mAP of 88.4%, which is notably higher than those of the State-of-the-Art algorithms. Moreover, under comparable conditions in terms of the number of parameters, FLOPs, and model size, YOLOv8l-Segment demonstrates superior detection accuracy compared to YOLOv5l-Segment, indicating that YOLOv8l-Segment serves as a more effective baseline network for the current task.

Table 3. Performance comparison of different models for seabed topography instance segmentation on the SG-MKD dataset.

Model	Params/M	FLOPs/s	Mask mAP50	Size/Mb
YOLOv5l-Segment	48.15	172.1	70.5	95.6
ASF-YOLO	47.20	165.0	80.6	170.0
Mask R-CNN	44.0	190.0	68.1	307.9
Cascade Mask R-CNN	77.1	1713	86.4	687.1
YOLOv8l-Segment	45.9	220.8	75.2	92.3
YOLO-SG	46.18	221.8	88.4	94.3

5. Discussion

Although deep-learning methods have shown great potential in the recognition and segmentation of seafloor topography and geomorphology, research in this field is still

relatively limited. This study constructed a seafloor topography and geomorphology dataset, SG-MKD, and proposed the YOLO-SG algorithm, achieving efficient recognition and segmentation of geomorphological units such as seamounts, sea knolls, and submarine depressions, advancing the application of deep learning in the recognition and delineation of small seafloor geomorphological units. The YOLO-SG algorithm proposed in this paper is based on a design concept of multi-scale and strong generalization ability, effectively addressing many challenges in seafloor topography recognition and segmentation. In the experimental results, YOLO-SG significantly improved target detection and segmentation accuracy compared with the YOLOv8l-Segment baseline network while also addressing the issue of imbalance in multi-scale object detection accuracy.

However, because of challenges in the instance segmentation dataset creation process, such as time-consuming labeling, complex annotations, and difficulty in maintaining consistent quality, the sample size and quality of the SG-MKD dataset still have significant room for improvement. In the future, it is hoped that more researchers with experience in underwater terrain and landform segmentation will participate in the expansion of the dataset, making its sample size and variety more diverse. In addition, we plan to introduce multi-source data and combine simulation sample generation techniques to address the issue of class imbalance in the dataset. This will help form a dataset with more diverse categories and a more balanced sample distribution, supporting the continuous optimization and broad application of this research.

6. Conclusions

In response to the practical needs of seabed terrain and landform recognition and segmentation, this study constructed a seabed terrain and landform instance segmentation dataset, SG-MKD, and proposed a seabed landform unit recognition and segmentation algorithm, YOLO-SG, based on lightweight upsampling operator and attention mechanisms.

The SG-MKD dataset contains 419 images with a resolution of 120 $\times$ 120, covering terrain and landform units such as seamounts, sea knolls, and submarine basins, providing important data support for seabed terrain and landform instance segmentation. YOLO-SG was designed with significant innovations, incorporating a feature extraction network that integrates the CBAM attention mechanism and a feature fusion network that introduces the CARAFE operator.

The feature extraction network, which integrates the CBAM attention mechanism, enhances the ability of the model to capture spatial information while focusing on key channel information. This allows the network to perform more balanced feature extraction of both local and global features, with a notable advantage in extracting the features of small targets. The feature fusion network, which incorporates the CARAFE operator, enhances the model's ability to focus on multi-scale targets via efficient multi-scale feature fusion and rich feature expression. This not only improves the model's attention to multi-scale objects but also effectively reduces interference from background noise.

The experimental results show that YOLO-SG demonstrates significant advantages in detecting and segmenting seabed topography features, with an average detection accuracy of 88.4%. Compared with the baseline network algorithm, the accuracy improved by 14.7% with a minimal increase in the computational complexity. In addition, the algorithm exhibits strong generalization capabilities, making it suitable for multiple regions and capable of performing seabed topography segmentation tasks across various input sizes.

In conclusion, by constructing the SG-MKD dataset and improving the YOLOv8l-Segment network, this study provides an efficient and reliable method for seabed topography instance segmentation.

Author Contributions: Conceptualization, Y.J., D.Z., X.Q., M.W. and Q.W.; Methodology, Y.J. and F.Y.; Resources, Y.J., Z.W., F.Y. and D.Z.; Supervision, Z.W.; Validation, Z.W. and F.Y.; Visualization, Y.J.; Writing—original draft, Y.J., Z.W. and F.Y.; Writing—review and editing, D.Z., X.Q., M.W. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (grant number 2022YFC2806600 and 2022YFC2806605), the Scientific Research Fund of the Second Institute of Oceanography, MNR (grant No. QNYC2403), the National Natural Science Foundation of China (grant numbers: 42306210), and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (grant numbers: SL2022ZD205, SL2023ZD102 and SL2023ZD203).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to acknowledge SIO (Scripps Institution of Oceanography) for providing the radar altimeter-derived gravity data and the bathymetry model Topo_27.1 used in this study; The IHO (International Hydrographic Organization) provided the Compilation Group (2023) GEBCO 2023 Grid.

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

References

1. Visbeck, M. Ocean Science Research Is Key for a Sustainable Future. *Nat. Commun.* **2018**, *9*, 690. [CrossRef] [PubMed]
2. Boetius, A.; Heilmayer, M. Mind the Seafloor. *Science* **2018**, *359*, 34–36. [CrossRef] [PubMed]
3. Danovaro, R.; Fanelli, E.; Aguzzi, J.; Billett, D.; Carugati, L.; Corinaldesi, C.; Dell'Anno, A.; Gjerde, K.; Jamieson, A.J.; Kark, S.; et al. Ecological Variables for Developing a Global Deep-Ocean Monitoring and Conservation Strategy. *Nat. Ecol. Evol.* **2020**, *4*, 181–192. [CrossRef]
4. Loureiro, G.; Dias, A.; Almeida, J.; Martins, A.; Hong, S.; Silva, E. A Survey of Seafloor Characterization and Mapping Techniques. *Remote Sens.* **2024**, *16*, 1163. [CrossRef]
5. Wu, Z.; Yang, F.; Tang, Y. *High-Resolution Seafloor Survey and Applications*; Science Press: Beijing, China, 2021.
6. Novaczek, E.; Devillers, R.; Edinger, E. Generating Higher Resolution Regional Seafloor Maps from Crowd-Sourced Bathymetry. *PLoS ONE* **2019**, *14*, e0216792. [CrossRef] [PubMed]
7. Torroba, I.; Bore, N.; Folkesson, J. A Comparison of Submap Registration Methods for Multibeam Bathymetric Mapping. In Proceedings of the 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV), Porto, Portugal, 6–9 November 2018; pp. 1–6.
8. Wang, X.; Yang, F.; Zhang, H.; Su, D.; Wang, Z.; Xu, F. Registration of Airborne LiDAR Bathymetry and Multibeam Echo Sounder Point Clouds. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
9. Harris, P.T.; Macmillan-Lawler, M.; Rupp, J.; Baker, E.K. Geomorphology of the Oceans. *Mar. Geol.* **2014**, *352*, 4–24. [CrossRef]
10. Lucieer, V.; Lamarche, G. Unsupervised Fuzzy Classification and Object-Based Image Analysis of Multibeam Data to Map Deep Water Substrates, Cook Strait, New Zealand. *Cont. Shelf Res.* **2011**, *31*, 1236–1247. [CrossRef]
11. Li, Z.; Peng, Z.; Zhang, Z.; Chu, Y.; Xu, C.; Yao, S.; García-Fernández, Á.F.; Zhu, X.; Yue, Y.; Levers, A.; et al. Exploring Modern Bathymetry: A Comprehensive Review of Data Acquisition Devices, Model Accuracy, and Interpolation Techniques for Enhanced Underwater Mapping. *Front. Mar. Sci.* **2023**, *10*, 1178845. [CrossRef]
12. Jung, J.; Lee, J.; Parrish, C.E. Inverse Histogram-Based Clustering Approach to Seafloor Segmentation from Bathymetric Lidar Data. *Remote Sens.* **2021**, *13*, 3665. [CrossRef]
13. Cutter, G.R.; Rzhanov, Y.; Mayer, L.A. Automated Segmentation of Seafloor Bathymetry from Multibeam Echosounder Data Using Local Fourier Histogram Texture Features. *J. Exp. Mar. Biol. Ecol.* **2003**, *285–286*, 355–370. [CrossRef]
14. Brown, C.J.; Blondel, P. Developments in the Application of Multibeam Sonar Backscatter for Seafloor Habitat Mapping. *Appl. Acoust.* **2009**, *70*, 1242–1247. [CrossRef]
15. Lawrence, E.; Hayes, K.R.; Lucieer, V.L.; Nichol, S.L.; Dambacher, J.M.; Hill, N.A.; Barrett, N.; Kool, J.; Siwabessy, J. Mapping Habitats and Developing Baselines in Offshore Marine Reserves with Little Prior Knowledge: A Critical Evaluation of a New Approach. *PLoS ONE* **2015**, *10*, e0141051. [CrossRef]

16. Chakraborty, B.; Menezes, A.; Dandapath, S.; Fernandes, W.A.; Karisiddaiah, S.M.; Haris, K.; Gokul, G.S. Application of Hybrid Techniques (Self-Organizing Map and Fuzzy Algorithm) Using Backscatter Data for Segmentation and Fine-Scale Roughness Characterization of Seepage-Related Seafloor Along the Western Continental Margin of India. *IEEE J. Ocean. Eng.* **2015**, *40*, 3–14. [CrossRef]
17. Lacharité, M.; Brown, C.J.; Gazzola, V. Multisource Multibeam Backscatter Data: Developing a Strategy for the Production of Benthic Habitat Maps Using Semi-Automated Seafloor Classification Methods. *Mar. Geophys. Res.* **2018**, *39*, 307–322. [CrossRef]
18. Pirtle, J.L.; Weber, T.C.; Wilson, C.D.; Rooper, C.N. Assessment of Trawlable and Untrawlable Seafloor Using Multibeam-Derived Metrics. *Methods Oceanogr.* **2015**, *12*, 18–35. [CrossRef]
19. Masetti, G.; Mayer, L.A.; Ward, L.G. A Bathymetry- and Reflectivity-Based Approach for Seafloor Segmentation. *Geosciences* **2018**, *8*, 14. [CrossRef]
20. Krizhevsky, A.; Krizhevsky, A.; Hinton, G. *Learning Multiple Layers of Features from Tiny Images*; Technical Report; University of Toronto: Toronto, ON, Canada, 2009.
21. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
22. Li, Z.H.; Cui, J.X.; Lu, H.P.; Zhou, F.; Diao, Y.L.; Li, Z.X. Prediction Model of Measurement Errors in Current Transformers Based on Deep Learning. *Rev. Sci. Instrum.* **2024**, *95*, 4. [CrossRef]
23. Li, Z.; Cui, J.; Chen, H.; Lu, H.; Zhou, F.; Rocha, P.R.; Yang, C. Research Progress of All-Fiber Optic Current Transformers in Novel Power Systems: A Review. *Microw. Opt. Technol. Lett.* **2025**, *67*, e70061. [CrossRef]
24. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo Algorithm Developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
25. Hafiz, A.M.; Bhat, G.M. A Survey on Instance Segmentation: State of the Art. *Int. J. Multimed. Info. Retr.* **2020**, *9*, 171–189. [CrossRef]
26. Yue, X.; Qi, K.; Na, X.; Zhang, Y.; Liu, Y.; Liu, C. Improved YOLOv8-Seg Network for Instance Segmentation of Healthy and Diseased Tomato Plants in the Growth Stage. *Agriculture* **2023**, *13*, 1643. [CrossRef]
27. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
28. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
29. Chen, B.; Huang, Y.; Xia, Q.; Zhang, Q. Nonlocal Spatial Attention Module for Image Classification. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1729881420938927. [CrossRef]
30. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-Aware Reassembly of Features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.
31. Saadat, H.; Bonnell, R.; Sharifi, F.; Mehuys, G.; Namdar, M.; Ale-Ebrahim, S. Landform Classification from a Digital Elevation Model and Satellite Imagery. *Geomorphology* **2008**, *100*, 453–464. [CrossRef]
32. Wilson, J.P. Digital Terrain Modeling. *Geomorphology* **2012**, *137*, 107–121. [CrossRef]
33. Reddy, G.P.O. Remote Sensing and GIS in Digital Terrain Modeling. In *Geospatial Technologies in Land Resources Mapping, Monitoring and Management*, 1st ed.; Reddy, G.P.O., Singh, S.K., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 201–222. ISBN 978-3-319-78711-4.
34. Liu, Y.; Wu, Z.; Zhao, D.; Zhou, J.; Shang, J.; Wang, M.; Zhu, C.; Luo, X. Construction of High-Resolution Bathymetric Dataset for the Mariana Trench. *IEEE Access* **2019**, *7*, 142441–142450. [CrossRef]
35. Libina, N.V.; Nikiforov, S.L. Digital Elevation Models of the Bottom in the Operational Oceanography System. *Oceanology* **2020**, *60*, 854–860. [CrossRef]
36. Tozer, B.; Sandwell, D.T.; Smith, W.H.; Olson, C.; Beale, J.R.; Wessel, P. Global Bathymetry and Topography at 15 Arc Sec: SRTM15+. *Earth Space Sci.* **2019**, *6*, 1847–1864. [CrossRef]
37. Sandwell, D.T.; Harper, H.; Tozer, B.; Smith, W.H.F. Gravity Field Recovery from Geodetic Altimeter Missions. *Adv. Space Res.* **2021**, *68*, 1059–1072. [CrossRef]
38. Gevorgian, J.; Sandwell, D.T.; Yu, Y.; Kim, S.S.; Wessel, P. Global Distribution and Morphology of Small Seamounts. *Earth Space Sci.* **2023**, *10*, e2022EA002331. [CrossRef]
39. Wessel, P.; Sandwell, D.T.; Kim, S.-S. The Global Seamount Census. *Oceanography* **2010**, *23*, 24–33. [CrossRef]
40. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
41. DeVries, T.; Taylor, G.W. Dataset Augmentation in Feature Space. *arXiv* **2017**, arXiv:1702.05538.
42. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved Techniques for Training GANs. In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016; pp. 2234–2242.

43. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
44. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. AutoAugment: Learning Augmentation Policies from Data. *arXiv* **2018**, arXiv:1805.09501.
45. Li, F.; Li, X.; Liu, Q.; Li, Z. Occlusion Handling and Multi-Scale Pedestrian Detection Based on Deep Learning: A Review. *IEEE Access* **2022**, *10*, 19937–19957. [CrossRef]
46. Jocher, G. YOLO by Ultralytics (version 5.7.0). GitHub. 2022. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 March 2025).
47. Kang, M.; Ting, C.M.; Ting, F.F.; Phan, R.C.W. ASF-YOLO: A Novel YOLO Model with Attentional Scale Sequence Fusion for Cell Instance Segmentation. *Image Vis. Comput.* **2024**, *147*, 105057. [CrossRef]
48. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6154–6162.

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.