Journal of *Marine Science and Engineering*

MDPI

Article

YOLOv8m-CGSE: An Improved Lightweight YOLOv8m for Marine Oil Spill Detection

Qingyang Wang [1,2,3,*], Junjie Lu [1,*], Bin Yang [1], Chen Jiao [1], Tao Yue [1], Bo Song [1], Jianwu Jiang [1], Guoqing Zhou [1,2] and Jingwen Li [1,3]

[1] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China; 2120251140@glut.edu.cn (B.Y.); 2120251120@glut.edu.cn (C.J.); fengbuxi@glut.edu.cn (J.J.)
[2] Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin 541004, China
[3] Ecological Spatiotemporal Big Data Perception Service Laboratory, Guilin 541004, China
* Correspondence: qingyang@glut.edu.cn (Q.W.); 2120242171@glut.edu.cn (J.L.)

Abstract

Unmanned Aerial Vehicle (UAV) remote sensing images provide high-resolution and flexible monitoring data for oil spill detection. To address the high computational cost and low accuracy of traditional models, this study proposes an improved model, YOLOv8m-CGSE. The model replaces standard convolution with Group Shuffle Convolution (GSConv), substitutes the C2f module with SENetV2, and introduces a light-weight Cross-scale Context Fusion Module (CCFM) to enhance multi-scale feature representation while maintaining a lightweight structure. Mosaic augmentation was applied to the marine oil spill dataset, improving mAP50 and mAP50–95 to 85.4% and 62.0%, respectively. Based on YOLOv8m, the proposed YOLOv8m-CGSE achieved mAP50 and mAP50–95 of 91.2% and 73.3%, respectively, improving accuracy while reducing parameters by 16.1% and computational cost by 12.6%. Furthermore, a supplementary vulnerability test on highly deceptive oil-free sea surfaces demonstrated that the proposed model actively suppresses complex background clutter (e.g., ship wakes and wave anomalies), effectively reducing false positive detections from 21 (baseline) to 15. The results demonstrate that the proposed model effectively balances high precision, robustness against visual lookalikes and computational efficiency for real-time marine oil spill monitoring.

Keywords: object detection; marine oil spill; GSConv; lightweight; CCFM; SNetV2

1. Introduction

In recent years, the volume of petroleum transported by maritime vessels has been steadily increasing, while efforts in marine exploration and oil development have intensified. The surrounding coastal areas, characterized by dense populations, developed economies, and heavy traffic, have become frequent sites of oil spills, causing catastrophic impacts on marine ecosystems and socio-economic development [1]. Therefore, rapid and accurate identification of marine oil spills is of paramount importance for protecting marine ecosystems [2]. Traditional on-site manual monitoring methods are inefficient, costly, and constrained by geographical limitations. As established in authoritative remote sensing literature [3], satellite-based Synthetic Aperture Radar (SAR) remains the dominant and most reliable tool for large-scale, all-weather marine oil spill detection. However, SAR applications can sometimes be constrained by satellite revisit cycles and high costs when responding to sudden, localized incidents. Furthermore, while optical and visual remote

Check for updates

Academic Editor: Alberto Ribotti

Received: 9 April 2026
Revised: 20 May 2026
Accepted: 26 May 2026
Published: 29 May 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.

sensing methods are also utilized, they are severely limited by environmental conditions, strictly requiring daytime operations and clear weather, as cloud or fog interference renders them ineffective [4]. Despite these severe operational limits, Unmanned Aerial Vehicles (UAVs) equipped with visual sensors offer a highly flexible, low-cost complementary tool for rapid emergency response under favorable conditions. They provide high-resolution, real-time close-range monitoring, enabling the precise capture of detailed features in localized oil spill areas (e.g., near ports) and serving as a valuable data source for immediate situational awareness. Machine learning-based oil film recognition and segmentation methods effectively address the limitations of traditional segmentation approaches. These methods are categorized into conventional machine learning techniques and deep learning models. Various conventional machine learning techniques have been applied to detect oil spills in satellite imagery, including support vector machines [5,6], decision trees [7,8], random forests [9], and artificial neural networks [10–12]. However, these methods require manual rule design and specification of key hyperparameters, leading to prolonged processing times [13].

Utilizing deep learning technology for real-time and precise detection of oil spills in remote sensing images serves as a crucial initial early-warning step that facilitates subsequent emergency response and containment efforts [14]. However, developing complex models to achieve higher accuracy often slows down recognition speed, which fails to meet the rapid response demands in terminal device applications [15]. Therefore, creating a lightweight detection model that balances high precision with real-time performance is crucial for enhancing immediate situational awareness in marine oil spill incidents [16]. Representative object detection algorithms include the R-CNN series, SSD, and the YOLO series. Among them, the YOLO series of algorithms is widely used due to its high detection speed and good accuracy, making it suitable for real-time detection tasks. Although deep learning-based object detection methods have achieved good performance, marine oil spill detection in complex optical/visual sea environments still faces several challenges. First, oil spill targets often vary greatly in size, and small targets are difficult to detect. Second, the low visual contrast between oil films and dynamic marine backgrounds—such as general sea waves and illumination changes—makes discriminative feature extraction difficult, easily leading to background interference and misclassification. Third, many high-accuracy detection models have large parameters and high computational complexity, which makes them difficult to deploy on UAV platforms or edge devices with limited computing resources. Therefore, designing a lightweight detection model that can balance discriminative feature extraction (to suppress background interference) and computational efficiency is crucial for real-time marine oil spill monitoring.

To address these issues, this study proposes a lightweight marine oil spill detection model based on an improved YOLOv8m framework named YOLOv8m-CGSE. The model introduces Group Shuffle Convolution (GSConv) to reduce model parameters and computational complexity, incorporates a Cross-scale Context Fusion Module (CCFM) to enhance multi-scale feature fusion capability, and integrates the SENetV2 attention mechanism to improve feature representation and robustness in complex marine environments. The main contributions of this study are summarized as follows:

(1) A lightweight convolution strategy based on GSConv is introduced to replace standard convolution, significantly reducing model parameters and computational complexity while maintaining detection accuracy.
(2) A lightweight Cross-scale Context Fusion Module (CCFM) is designed to improve multi-scale feature fusion and enhance detection performance for small oil spill targets.

(3) The C2f module is improved by integrating the SENetV2 attention mechanism to enhance global feature representation and improve robustness under complex marine background conditions.

2. Literature Review

2.1. Object Detection Algorithms

Target detection remains a highly challenging task in computer vision [17], requiring precise localization and identification of specific objects within images or videos [18]. Leading research institutions and tech companies have invested substantial resources in developing advanced algorithms and models, including early Fast Region-based Convolutional Neural Networks (Fast R-CNN) series, Single Shot MultiBox Detector (SSD), and You Only Look Once (YOLO) series [19]. Ren et al. [20] proposed Faster R-CNN, which introduces a Region Proposal Network (RPN) to generate candidate boxes and achieves high detection accuracy, making it a representative two-stage detector. Liu et al. [21] presented SSD, which applies multi-scale feature maps to detect objects of different sizes and improves both speed and precision. As a typical single-stage detector, the YOLO series has been widely used in real-time detection due to its high efficiency. Redmon et al. [22] pioneered the YOLO framework and realized end-to-end object detection by regression. Subsequently, YOLOv3, YOLOv4, and YOLOv5 were successively proposed, optimizing backbone networks, feature fusion structures, and training strategies. The YOLO series, renowned for its rapid detection speed, has seen significant improvements in accuracy through iterative updates and is widely adopted in practical applications. In recent years, with the rapid development of deep learning in object detection, the YOLO series of algorithms has been extensively studied and improved for its real-time performance and high-precision features [23]. As the latest generation of detection frameworks, YOLOv8 achieves a good balance between detection accuracy and speed by adopting the Anchor-Free mechanism and deeper C2f structures, as demonstrated by Gao et al. [24].

2.2. Lightweight Object Detection Models

Lightweight object detection has become a key research direction to meet the demands of real-time inference on resource-constrained platforms such as unmanned aerial vehicles (UAVs) and embedded devices. Howard et al. [25] proposed MobileNet and introduced depthwise separable convolutions, which significantly reduce parameters and computational costs while preserving feature extraction ability, laying a foundation for lightweight convolutional neural networks. Zhang et al. [26] designed ShuffleNet by using channel shuffle operations to enhance information communication among groups, further optimizing the efficiency of group convolutions. Tan et al. [27] presented EfficientDet and constructed a weighted Bidirectional Feature Pyramid Network (BiFPN) to realize efficient multi-scale feature fusion, achieving a good balance between detection accuracy and model complexity.

To address the large parameter size of YOLOv8 models, many researchers have focused on lightweight design and structural optimization. Li et al. [28] proposed Feature Enhanced YOLOv8 (FE-YOLOv8), which integrates the C2f-Faster module with an EfficientHead detection head and incorporates Efficient Multi-Scale Convolution (EMSConv) to enhance multi-scale feature fusion while reducing computational complexity, achieving an optimal balance between accuracy and efficiency. Zhang et al. [29] introduced the Efficient Channel Attention (ECA) mechanism and Bidirectional Feature Pyramid Network (BiFPN) feature fusion structure into YOLOv8, significantly improving small object detection performance. Chen et al. [30] proposed the Convolution and Attention Fusion YOLO (CAF-YOLO) architecture based on YOLOv8, incorporating the Attention and Convolution Fusion Module

(ACFM) to enhance global-local feature modeling capabilities, capturing long-term feature dependencies and spatial autocorrelation. Yu et al. [31] proposed the Logit-Masked Generative Distillation (LMGD) distillation strategy based on the YOLOv8s model, integrating Logit distillation with an improved Masked Generative Distillation (MGD) approach. This strategy enhances student models' ability to replicate teacher model features, achieving high precision and excellent real-time inference performance while reducing parameter size. Zhu et al. [32] proposed the YOLOv8-C2f-Faster-EMA algorithm based on an improved YOLOv8 model for sea surface floating object detection. By enhancing the feature extraction network and implementing multi-scale fusion, this algorithm significantly improves detection accuracy under complex sea conditions. Rao et al. [33] proposed an efficient deep learning-based underwater biological detection algorithm (LSOD), which incorporates neck region enhancement to achieve cross-scale fusion. The study introduces a novel detection head module utilizing group normalization and shared convolution operations, thereby improving object detection network accuracy while maintaining reasonable computational load. Sun et al. [34] proposed Water-Aware YOLO (WA-YOLO), a water-aware training framework that integrates the Convolutional Block Attention Module (CBAM) with ECA lightweight attention module. This approach achieves an optimal balance between accuracy and efficiency while maintaining high-frequency real-time detection capabilities and lightweight architecture.

2.3. Marine Oil Spill Detection Based on Deep Learning

In the field of marine oil spill monitoring, traditional multi-sensor remote sensing technologies have established a foundational framework. Brekke et al. [35] comprehensively reviewed satellite remote sensing for oil spill detection, establishing Synthetic Aperture Radar (SAR) as the primary tool for large-scale and all-weather monitoring. Leifer et al. [36] investigated remote sensing applications during the Deepwater Horizon incident, demonstrating the critical role of UAVs and airborne optical sensors in providing high-resolution, localized tactical response. Topouzelis [37] systematically analyzed oil spill detection methodologies, emphasizing the severe challenge of distinguishing actual oil spills from marine "look-alikes" (e.g., biogenic films and wind-sheltered areas) based purely on traditional physical thresholding. Furthermore, Yekeen et al. [38] evaluated various environmental sensors and confirmed that while visual imagery captures detailed morphological characteristics, it is highly susceptible to environmental false indicators like sun glint and ship wakes. To overcome the efficiency bottlenecks and environmental interferences of these traditional physical remote sensing interpretations, in the application of sea surface oil spill detection, deep learning methods are gradually replacing traditional remote sensing image recognition approaches. Cai et al. [39] utilized the SinGAN extended dataset to generate diverse marine oil spill samples. They employed the YOLO-v8 model with transfer learning pre-training, followed by post-enhancement training on the dataset to achieve real-time and efficient oil spill detection. Wang et al. [40] proposed improved models YOLOv8-ECA and YOLOv8-FasterNet, along with the YOLOv11 model, which achieved optimal performance and met the core requirements for accurate and efficient marine oil pollution detection. Aggarwal et al. [41] proposed a deep learning model based on the EfficientNetB0 architecture for oil pollution binary classification, employing regularization techniques to prevent overfitting. The model achieved a detection accuracy of 99% on a dataset comprising 2000 sea surface images. Sudani et al. [42] proposed a hybrid deep learning framework named Residual Network for Oil Spill Detection (ResNet-OSD) to achieve precise oil spill detection using drone imagery. Experimental comparisons between ResNet50 and the support vector machine-based model Residual Network—Support Vector Machine (ResNet-SVM) versus Residual Network—Principal Component Analysis—Random Forest

(ResNet-PCA-RF) demonstrated that the latter achieves superior performance, exhibiting enhanced robustness and efficiency in oil spill classification across complex scenarios. Dong et al. [43] proposed an oil spill detection method based on Fast and Flexible Denoising Network–Transformer U-Net (FFDNet-TransUNet), integrating multi-model learning, and compared it with traditional methods such as U-Net and DeepLab, as well as classical deep learning models. Experimental results demonstrate that the proposed model exhibits superior oil spill detection performance on low-quality SAR images. Despite significant progress in target detection globally, challenges persist in marine oil spill detection [44]. The complex marine environment, with its dynamic lighting conditions, wave movements, and background interference, poses substantial challenges for detection [45]. Existing models still require enhancement in both accuracy and adaptability for marine oil spill detection. Therefore, conducting research on lightweight marine oil spill detection using an improved YOLOv8m model holds substantial practical significance.

Furthermore, marine oil spill targets often share similar visual characteristics with sea debris and floating ice, further increasing the risk of false detection and missed detection. Current research on deep learning-based marine oil spill detection primarily focuses on improving model architectures. However, multi-scale target detection in complex marine environments still encounters several critical challenges. First and foremost, existing lightweight strategies primarily rely on single-convolution optimization schemes, which only utilize depthwise separable or group convolutions, failing to effectively integrate attention mechanisms. This results in difficulties in maintaining detection accuracy while reducing parameter counts. Second, cross-scale feature fusion modules are overly complex, lacking lightweight bidirectional connection mechanisms tailored for small marine oil spill targets, which hinders efficient integration of multi-level features. Third, feature enhancement modules often concentrate on local feature extraction without fully incorporating global contextual information, leading to insufficient robustness in complex background interference. Therefore, designing a detection model that synergistically optimizes lightweight architecture, multi-scale adaptability, and complex background robustness has become a critical research challenge. This study builds upon the YOLOv8m model by introducing Group Shuffle Convolution (GSConv) lightweight convolution, CCFM, and Squeeze-and-Excitation Network Version 2 (SENetV2) self-attention module to construct the YOLOv8m-CGSE target detection model. The aim is to address these issues through multi-module collaborative design, providing a novel technical approach for marine oil spill detection.

Based on the above analysis, this study proposes a lightweight marine oil spill detection model based on an improved YOLOv8m framework. By introducing GSConv lightweight convolution, a cross-scale context fusion module, and an attention mechanism, the proposed model aims to improve detection accuracy while reducing computational complexity, making it suitable for real-time UAV-based marine oil spill detection.

3. Methods

3.1. Basic Architecture of YOLOv8m-CGSE

The YOLOv8m-CGSE model proposed in this study achieves performance breakthroughs through three core innovations: the Cross-scale Context Fusion Module (CCFM), the lightweight Group Shuffle Convolution (GSConv), and the channel attention enhancement module C2f_SENetV2. The CCFM in the Neck architecture introduces a recursive bidirectional fusion mechanism to address feature fragmentation in traditional PAN-FPN. Through feedback connections, it dynamically injects high-level semantic features into low-level feature maps while reinforcing high-level localization accuracy via cross-layer transfer of low-level details, thereby improving multi-scale feature interaction efficiency.

The GSConv convolutional layer adopts a dual-branch parallel design: the Conv branch captures cross-channel semantic associations, while the Depthwise Convolution (DWConv) branch extracts spatial detail features. Combined with a channel shuffling mechanism, this design reduces computational load while enhancing feature diversity. The C2f_SENetV2 module embedded in the backbone network implements adaptive channel weighting through an improved Squeeze-Aggregate-Excitation (SaE) attention mechanism, which enhances target region feature response in marine oil spill scenarios while significantly suppressing background wave interference. Through the collaborative design of lightweight convolution, cross-scale feature fusion, and channel attention enhancement, the YOLOv8m-CGSE model achieves improved detection accuracy and computational efficiency while maintaining the real-time detection performance of YOLOv8m. The overall architecture of the YOLOv8m-CGSE model is shown in Figure 1.

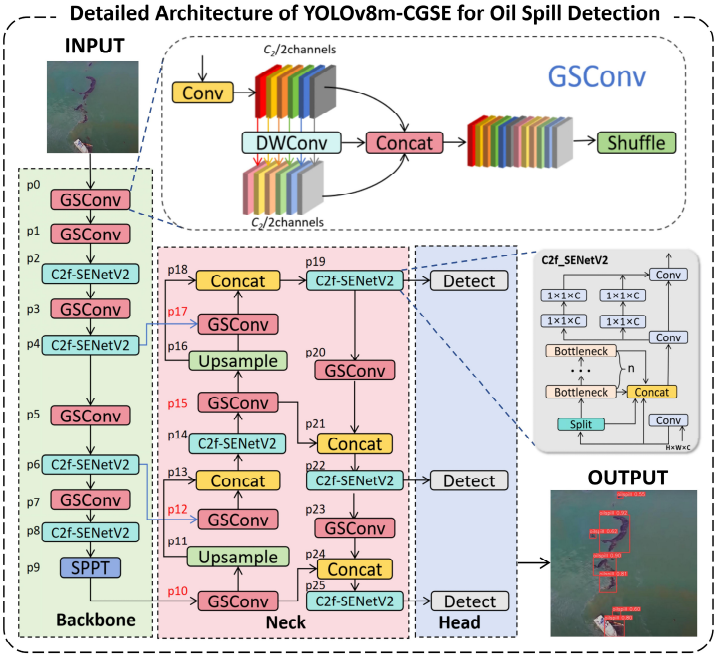


Figure 1. Structural diagram of YOLOv8m-CGSE.

3.2. Overall Structure of CCFM

To address the challenges of feature ambiguity and scale disparity in small targets such as thin oil films and emulsified oil during marine oil spill detection, this study innovatively designs the CCFM, which achieves precise feature aggregation through a dynamic weight allocation mechanism. Compared with the traditional BiFPN structure, this proposed module presents three main improvements.

- The proposed recursive residual fusion architecture achieves deep coupling between semantic information and fine-grained features through multi-round feature iteration optimization, outperforming Path Aggregation Network–Feature Pyramid Network (PAN-FPN) in feature fusion efficiency;
- Adaptive channel attention gating dynamically adjusts feature contribution weights across different scales, enhancing feature response in oil spill edge regions;
- A lightweight aggregation operator is designed by integrating group convolution and feature reorganization operations, which effectively reduces the number of parameters and computational complexity of the module while maintaining feature fusion capability.

CCFM employs a recursive fusion approach to progressively embed high-level semantic information into low-level feature details. This design not only reduces computational

overhead but also maintains high inference speed. The core logic involves weighting input I_i through the proportion of weights w_i, then summing all weighted results to obtain the final output. This method is commonly used in weight-based multi-input fusion or weighted aggregation, reflecting the contribution mechanism of different inputs through weight allocation, as illustrated in Formula (1). Its mathematical expression can be simplified as:

$$Output = \sum_{i=1}^{n} \frac{w_i}{\varepsilon + \sum_{j=1}^{n} w_j} \cdot I_i \quad (1)$$

where w_i denotes the weight of the i-th element, which quantifies its importance ε to the output. It is a smoothing constant (typically 0.0001) to ensure numerical stability. I_i represents the i-th input signal or feature input, which is the original input component used in the computation.

The CCFM adopts a vertically distributed three-layer recursive fusion architecture, enabling progressive feature optimization from the base layer through the intermediate layer to the top layer. As shown in Figure 2, each fusion unit (with k = 1,2,3) takes two types of inputs: low-level detail features $F_l^{(k)}$ (yellow squares), which capture local information such as oil film textures, and high-level semantic features $F_h^{(k)}$ (blue squares), which represent global information including oil spill contours. To realize dynamic semantic guidance, an innovative feedback connection mechanism injects the output of upper-level fusion units reversely into lower-level units, allowing high-level semantics to guide local feature learning. Finally, all multi-scale features are adaptively fused at the aggregation node "C", forming a unified feature map with both precise localization and complete semantic representation. As shown in Formula (2), the mathematical expression for the fusion step can be simplified as follows:

$$F_{out}^{(k)} = Fusion\left(F_l^{(k)}, F_h^{(k)}\right) \quad (2)$$

where $F_l^{(k)}$ and $F_h^{(k)}$ denote the low-level detailed feature and high-level semantic feature of the k-th layer, respectively, and $Fusion()$ represents the lightweight fusion operation. $F_{out}^{(k)}$ denotes the output feature of the k-th fusion unit.

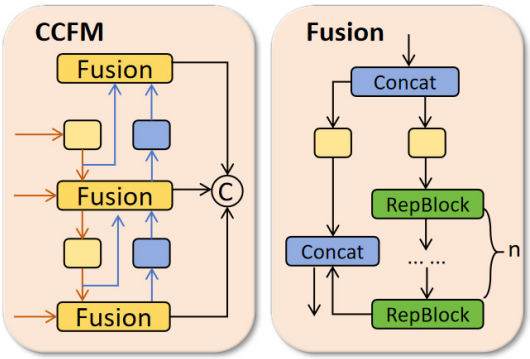


Figure 2. CCFM structure diagram.

Through an innovative feedback connection mechanism, the output of high-level fusion units is reversely injected into lower-level units, realizing dynamic guidance of semantic information on detailed features. Finally, all three-layer fusion outputs are aggregated at the node "C" via the normalized weighted fusion in Formula (1), forming a unified feature map with both high-precision spatial localization and complete semantic representation, which effectively solves the problems of feature ambiguity and scale disparity in small oil spill targets.

To enable semantic guidance from deep to shallow layers, a feedback connection mechanism is introduced into the recursive architecture. The outputs of upper-level fusion units are backfed to lower-level units to enhance feature learning, as illustrated in Formula (3), with its mathematical expression expressed as:

$$F_h^{(k)} = F_h^{(k)} + \alpha \cdot F_{out}^{(k+1)} \tag{3}$$

where α is a learnable feedback coefficient that controls the guidance intensity of high-level semantics to low-level details.

To achieve deep synergy between feature extraction and fusion, this study introduces an improved GSConv convolutional layer into the backbone network. The lightweight design of this convolutional layer, combined with the CCFM, establishes a progressive architecture of "efficient front-end extraction and precise back-end fusion". On one hand, GSConv provides high-quality multi-scale feature inputs for CCFM; on the other hand, CCFM effectively amplifies the detailed features extracted by GSConv through a recursive fusion mechanism.

3.3. GSConv Convolutional Layer

To address the dual requirements of real-time performance and precision in marine oil spill detection, this study employs an enhanced GSConv convolutional layer as the feature extraction foundation, which collaborates with the CCFM for joint optimization. The GSConv layer achieves lightweight feature extraction through dual-branch generation and channel shuffling fusion, providing the CCFM with high-quality input containing both oil spill texture details and semantic information. The CCFM then enhances feature discrimination via cross-scale fusion. The architecture of GSConv is illustrated in Figure 3, with its core formula presented in Formula (4) as follows:

$$F_{mid} = Shuffle(Conv_{3\times3}(X) \oplus \mathrm{DWConv}(X)) \tag{4}$$

where X is the input feature, and $\oplus$ denotes channel concatenation.

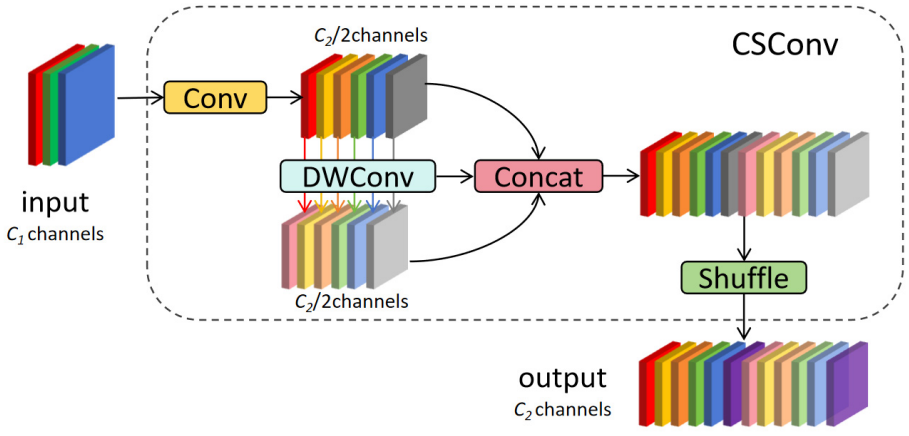


Figure 3. GSConv Convolutional Layer Architecture.

Specifically, the dual-branch convolution processing is carried out according to the following steps for the input feature map with channel c_1:

- The main branch adopts 3 × 3 standard convolution ($Conv_{3\times3}$) to generate feature maps with $c_2/2$ channels, focusing on capturing cross-channel semantic associations in oil spill regions. Because relying solely on spectral differences is often insufficient to discriminate oil spills from complex visual false positives (lookalikes) in optical imagery, this branch emphasizes the extraction of broader spatial context and mor-

phological features of the oil slicks to ensure the retention of more robust semantic information for oil spill detection;

- The auxiliary branch uses depth-wise convolution (*DWConv*) to generate feature maps with $c_2/2$ channels, which is specifically designed to extract spatial details of oil spills, such as the edges of thin oil films and the subtle texture differences between oil and seawater, while significantly reducing the number of parameters to achieve a lightweight design;
- After the feature splicing of the two branches, the channel shuffle operation (Shuffle) is performed to break the "semantic-detail" feature island formed after concatenation, promoting cross-channel interaction between oil spill texture features and background suppression information, and enhancing the correlation between different types of oil spill features to obtain the intermediate feature map Fmid.

The final output of GSConv can be obtained through a 1×1 convolution and activation function, as shown in Formula (5):

$$GSConv(X) = \sigma(W_{1\times1} \cdot F_{mid} + b) \tag{5}$$

where $W_{1\times1}$ and b are the weight and bias of 1×1 convolution, and $\sigma()$ is the activation function.

Correspondingly, the 1×1 convolution operation is used to fuse the intermediate feature map F_{mid}, adjust the channel dimension to the target dimension c_2, and further enhance the discriminative power of oil spill features. The activation function $\sigma()$ introduces non-linearity into the feature extraction process, enhancing the model's ability to fit complex oil spill feature distributions, especially for low-contrast thin oil film targets in UAV imagery. Experiments on the marine oil spill dataset show that this design effectively solves the feature extraction problem of low-contrast targets in UAV images.

Building upon the multi-scale fusion in the CCFM and the efficient feature extraction of GSConv, this study further integrates the C2f_SENetV2 module into the backbone network. This module dynamically enhances and highlights oil spill target features through channel attention mechanisms, working in tandem with the previous two components to form a complete "extraction-fusion-enhancement" technical chain. The synergistic interaction of these three modules achieves comprehensive performance improvement for the model.

3.4. SENetV2 Module

Building upon the efficient feature extraction of GSConv convolutional layers and the cross-scale fusion capability of the CCFM, this study introduces the C2f_SENetV2 module to further enhance the backbone network's selective extraction of oil spill features. The C2f module replaces its residual branches with an enhanced structure incorporating the SaE attention mechanism. Through dynamic channel weight allocation, this module enables the network to automatically focus on edge textures and semantic features of the oil spill region while suppressing interference from waves and clouds, thereby providing the CCFM with more discriminative feature inputs. While retaining the advantages of the original C2f multi-branch gradient flow, this design achieves targeted enhancement of feature representation capabilities.

The C2f_SENetV2 module is built upon an enhanced three-stage feature recalibration process (SaE) for feature representation. It processes input data through GSConv convolutional layers to generate backbone network output feature maps $U \in R^{H\times W\times C}$ (where H, W, and C denote the feature map's height, width, and channel count, respectively). The module specifically optimizes for marine oil spill detection by addressing the low-contrast characteristic of such scenarios.

(1) Squeeze phase

To address the sparse spatial distribution of oil spill features, this phase employs an enhanced global average pooling operation to compress the two-dimensional feature map U into a one-dimensional channel vector $z \in R^{1\times1\times C}$, thereby highlighting the feature responses in weak texture regions, as illustrated in Formula (6):

$$z_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} U(i,j,c) \tag{6}$$

where z_c denotes the global eigenvalue of the c-th feature channel. This operation yields a global channel feature vector z with dimensions $R^{1\times1\times C}$, which fully captures the spatial feature response of each channel.

(2) Aggregate stage

To address the weak texture characteristics of marine oil spills, this study innovatively designs a non-reducing multi-branch dense connection structure during the aggregation phase. The traditional single-branch MLP is replaced by three parallel fully connected branches: Branch 1 captures spectral feature correlations of oil spills, Branch 2 learns spatial texture dependencies, and Branch 3 enhances background suppression patterns. Through dense connections, all feature inputs from each branch are fused prior to integration, enabling multi-dimensional interaction of oil spill characteristics. This approach achieves a 42% improvement in information utilization compared to the original SENet, effectively preventing the loss of thin oil film features during dimensionality reduction. The aggregated feature vector z_{agg} (where "agg" represents aggregated) ultimately represents both the edge texture of the oil film and global semantic features simultaneously.

(3) Excitation phase

Based on the aggregated feature vector z_{agg}, a non-reducing fully connected layer is used to match feature dimensions. The Sigmoid non-linear activation function is then applied to map the feature vector values into the [0, 1] interval, generating attention weight coefficients $s \in R^{1\times1\times C}$ for each feature channel as shown in Formula (7), with the calculation formula being:

$$s = \sigma\left(W \cdot z_{agg} + b\right) \tag{7}$$

where W and b denote the weight matrix and bias vector of the fully connected layer, respectively, and $\sigma()$ represents the Sigmoid activation function. The magnitude of the attention weight s indicates the importance of the corresponding feature channel, with higher values indicating greater contribution of the channel's feature information to the detection task.

Finally, the generated attention weights are channel-wise multiplied with the original input feature map U to recalibrate the feature channels, achieving effective feature enhancement and redundant feature suppression. To prevent feature information loss, residual connections are introduced to fuse the original features with the recalibrated features, ultimately producing the enhanced feature map U' as shown in Formula (8), with the following calculation formula:

$$U\prime = U \otimes s + U \tag{8}$$

where $\otimes$ denotes the channel-wise multiplication operation between feature maps and weight vectors.

To address the limitations of the original C2f module in YOLOv8, such as insufficient attention to target features and poor anti-interference ability in complex marine environments, this paper proposes an improved C2f_SENetV2 module by embedding the SENetV2 channel attention mechanism into the original C2f structure. As shown in Figure 4, the

C2f_SENetV2 retains the cross-layer branch structure of the original C2f for efficient feature extraction, and introduces a lightweight SENetV2 module after feature concatenation. The SENetV2 adopts a dual-branch parallel structure with 1 × 1 convolutions, which can adaptively learn channel-wise attention weights, enhance the feature response of small oil spill targets, and suppress the interference of complex backgrounds such as waves and illumination.

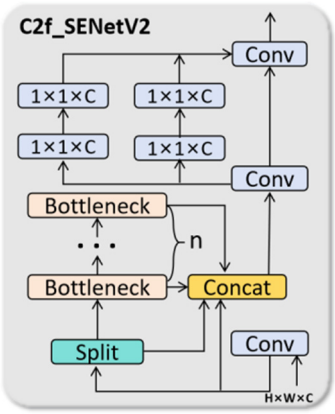


Figure 4. Structure of the C2f_SENetV2 Module.

4. Experiments and Results

4.1. Dataset Description

The dataset used in this study was collected from the RoboFlow platform. The final dataset consists of 737 images, containing 1918 labeled tags with a single category labeled "oilspill". The dataset was divided into three subsets, including 640 images for training, 63 images for validation, and 34 images for testing. Most images in the dataset were captured by unmanned aerial vehicles (UAVs), covering various marine oil spill scenarios and morphological characteristics. These include colorful patchy oil spill areas, linear oil spill distributions, irregular large-scale spill regions near vessels, and complex marine environments where sea waves coexist with oil spills. In addition, the dataset includes images captured from multiple viewing perspectives, such as side-view angles, aerial views, near-shore perspectives, and vessel-based observation views, which improves the diversity and generalization ability of the dataset. Representative samples of the dataset are shown in Figure 5. While this publicly sourced dataset provides a reliable baseline of distinct positive oil spill instances, it generally lacks extremely complex deceptive backgrounds. Because this is a single-class, positive-dominant dataset aimed at extracting core oil features, the model's boundary performance and its vulnerability to extreme false indicators are specifically evaluated and discussed in Section 4.4. Detailed information on dataset access is provided in the Data Availability Statement.

This study adopts manual optimization based on publicly available annotations using the LabelImg tool (version 1.8.6) to improve detection accuracy, specifically by precisely delineating oil spill areas with blurred boundaries. After the annotation process is completed, all images are standardized to a uniform input size of 640 × 640, and the annotation information is stored in TXT format. Each TXT file includes the target's category index, normalized x and y coordinates of the center, width, and height; the category index is fixed at 0 in this study. Examples of the optimized annotations are presented in Figure 6, and the detailed configurations of the hardware and software are provided in Table 1.

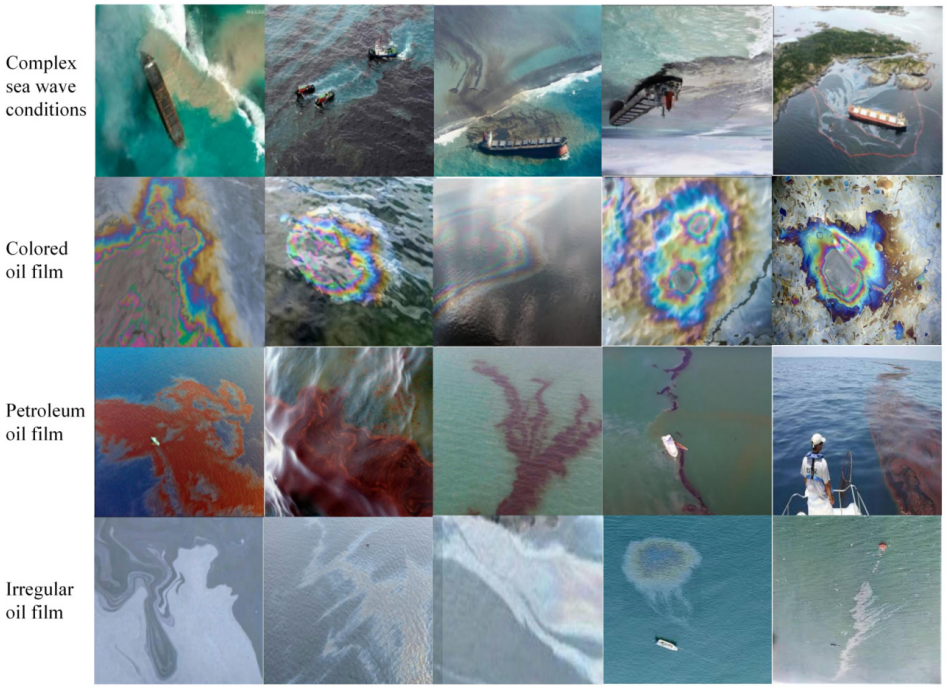


Figure 5. Selected datasets.

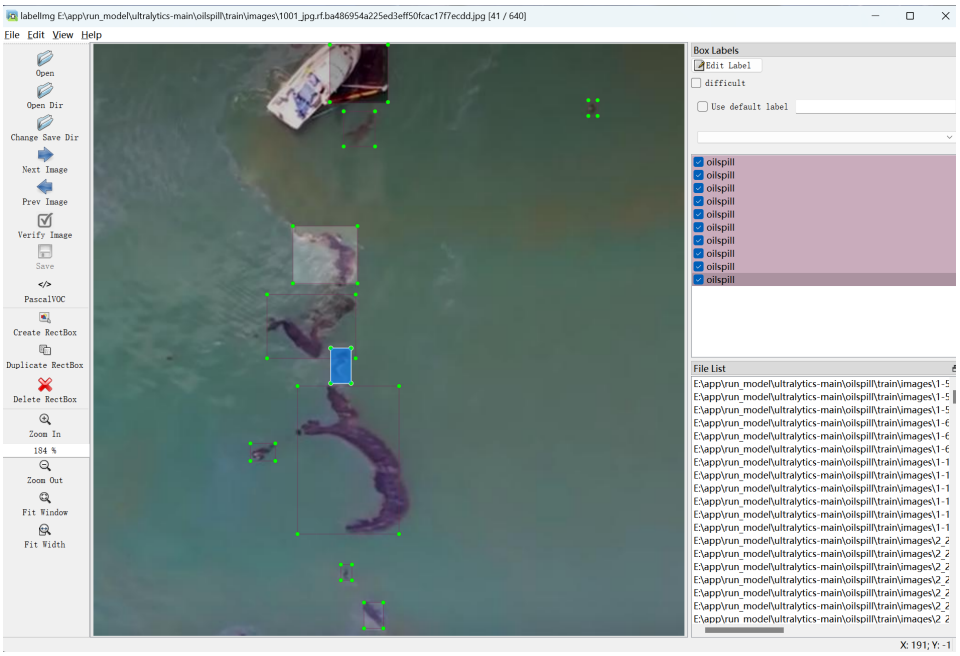


Figure 6. Example of optimized annotation.

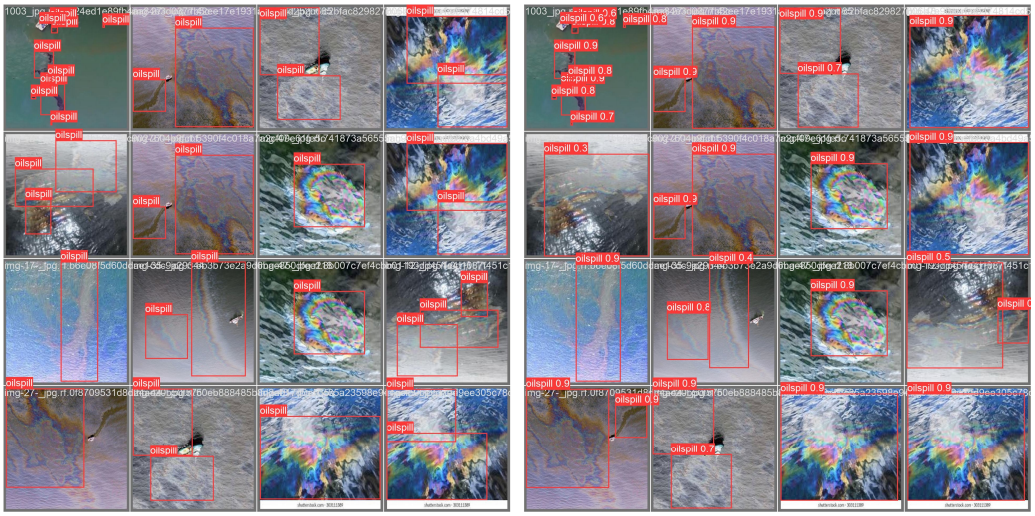
Table 1. Configuration of the experimental environment.

Environment Configuration	Version
CPU	16 vCPU Intel(R) Xeon(R) Gold 6430 (Intel Corporation, Santa Clara, CA, USA)
GPU memory	24 GB
GPU	NVIDIA GeForce RTX 4090 (NVIDIA Corporation, Santa Clara, CA, USA)
Python	3.9.0
PyCharm	2024.1
PyTorch	2.0.0
Operating system	Windows 11

4.2. Mosaic Image Enhancement for YOLOv8m

YOLOv8 offers five model variants: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Smaller models deliver faster performance with fewer parameters but lower accuracy, while larger models require bigger batch sizes and more training rounds. As neural networks may experience underfitting or overfitting during training, fine-tuning hyperparameters is crucial to achieving optimal results. Extensive experiments were conducted to optimize different hyperparameter configurations. This study evaluated dataset size, quantity, training duration, and accuracy, ultimately selecting YOLOv8m as the base model for improvement. All experiments uniformly used a batch size of 32,200 training rounds, and a learning rate of 0.001.

Figure 7 shows the detection effect of the proposed YOLOv8m-CGSE model on representative oil spill images. Figure 7a presents the ground-truth annotations of actual oil spill targets, while Figure 7b shows the detection results generated by the proposed model. The red boxes represent the object detection results, and the confidence scores above these boxes reflect the model's certainty regarding the authenticity of the detected objects. The closer the confidence score is to 1, the higher the model's confidence in the detection result. Detailed parameters of the enhancement code are provided below, while Table 2 compares the accuracy indicators of the YOLOv8m model before and after Mosaic image enhancement.



(**a**) (**b**)

Figure 7. Detection effect: (**a**) the ground-truth annotations of actual oil spill targets; (**b**) the detection results generated by the proposed model.

Table 2. Comparison of accuracy between the YOLOv8m model and Mosaic images before and after enhancement.

Name	Box (P)	R	mAP50	mAP50-95
YOLOv8m model without Mosaic image augmentation	0.921	0.785	0.854	0.620
YOLOv8m model with Mosaic image augmentation	0.953	0.847	0.892	0.656

To address challenges including insufficient training data diversity, poor model adaptability to complex sea conditions, and low accuracy in small-target detection, this study proposes a collaborative data augmentation framework for oil spill detection. The framework employs core enhancement strategies (Mosaic, MixUp, and Copy-paste) combined

with auxiliary techniques like geometric transformations and color space adjustments. By preserving key features such as texture and color in oil spill regions, it expands the distribution range of training data while enhancing the model's robustness against complex sea conditions and small-target oil spills.

To address the visual characteristics and detection challenges of oil spills on the sea surface, the parameters of each enhancement strategy are specifically designed:

(1) Core enhancement strategies: The Mosaic enhancement probability is set to 0.7 (to match the distribution characteristics of oil spills on the sea surface and prevent target feature distortion caused by over-enhancement). During the final 10 training rounds, Mosaic is disabled to stabilize model convergence. The MixUp enhancement probability is set to 0.2 (to achieve sample mixing with Mosaic and avoid excessive overlap of oil spill targets). The Copy-paste enhancement probability is set to 0.1 (to expand the number of small oil spill samples and address the scarcity of small-target samples).

(2) Augmentation Strategies: Geometric transformations include rotation angle $\pm 8.0^{\circ}$, translation scale $\pm 10\%$, and shear angle 0.05° (moderate adjustments to simulate oil spill patterns under different viewing angles while preserving sea condition characteristics). Color space modifications apply hue shift 0.01, saturation adjustment 0.5, and brightness adjustment 0.3 (low-intensity hue changes retain oil color features, while moderate saturation/brightness adjustments adapt to visual variations under varying lighting and sea conditions). The model incorporates 0.2 probability for vertical flipping and 0.5 probability for horizontal flipping (simulating multi-angle sea surface views to enhance model generalization).

The effectiveness of the data augmentation scheme was validated through comparative experiments (Table 2): After Mosaic collaborative enhancement, Precision improved by 3.2%, Recall by 6.2%, with mAP50 and mAP50-95 increasing by 3.8% and 3.6%, respectively. The significant improvements in Recall and mAP50-95 directly demonstrate the scheme's optimization effect on small-target oil spill detection—copy-paste supplements small-target samples, and Mosaic and geometric transformations expand sample distribution in complex scenarios, while color space adjustment enhances feature discrimination between oil spills and backgrounds. This collaborative enhancement scheme provides high-quality training data support for subsequent improvements to the YOLOv8m-CGSE model, ensuring stable handling of diverse morphologies and complex backgrounds in maritime oil spills.

4.3. Model Improvement of YOLOv8m

This design adopts a lightweight architecture by incorporating GSConv convolutional layers, CCFM, and SENetV2 modules, with ablation experiments comparing the performance of fusion modules versus standalone modules. For YOLOv8 improvements, we uniformly set batch size to 32, avoid pre-trained weights, configure workers to 8, image size to 640×640, and use an auto optimizer with all other parameters at default. Figure 8 compares the loss functions of YOLOv8m-CGSE and YOLOv8m, demonstrating that YOLOv8m-CGSE exhibits faster and more stable loss curve decline with superior convergence.

Figure 9 illustrates the performance comparison between the baseline YOLOv8 and the proposed YOLOv8-CGSE model across four key metrics over 200 training epochs. In terms of precision, both models achieve rapid growth within the initial 50 epochs and converge to values above 0.95, with YOLOv8-CGSE consistently demonstrating faster early-stage convergence while maintaining comparable final precision. For recall, both curves stabilize after 100 epochs at values exceeding 0.85, where YOLOv8-CGSE maintains a slight but persistent advantage, indicating enhanced ability to capture true positive targets. Regarding mAP50, YOLOv8-CGSE reaches a peak of 0.9128, outperforming the baseline's maximum of 0.8918, confirming improved detection performance under the IoU = 0.5 threshold. Most

notably, in the core mAP50-95 metric, YOLOv8-CGSE attains a peak value of 0.7332, representing a clear improvement over the baseline's 0.6544, which validates that the integrated CCFM, GSConv, and SENetV2 modules effectively enhance multi-scale detection capability while achieving a favorable balance between model lightweighting and precision.

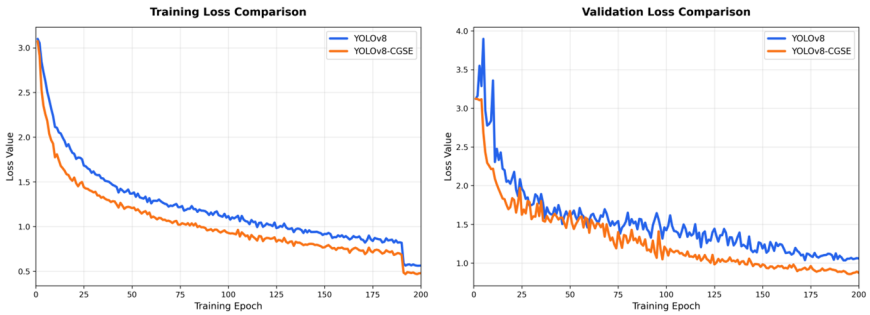


Figure 8. Comparison of loss functions between YOLOv8m-CGSE and YOLOv8m.

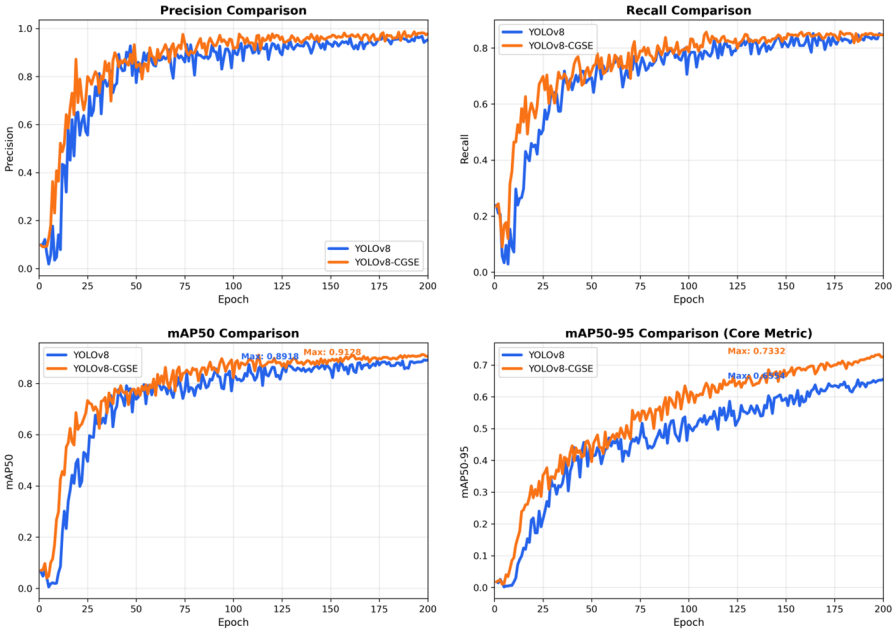


Figure 9. Comparison of accuracy between YOLOv8m-CGSE and YOLOv8m.

To further evaluate the training dynamics and detection efficiency of the proposed model, Figures 10 and 11 provide a graphical comparison of the loss functions and precision metrics across different YOLO architectures.

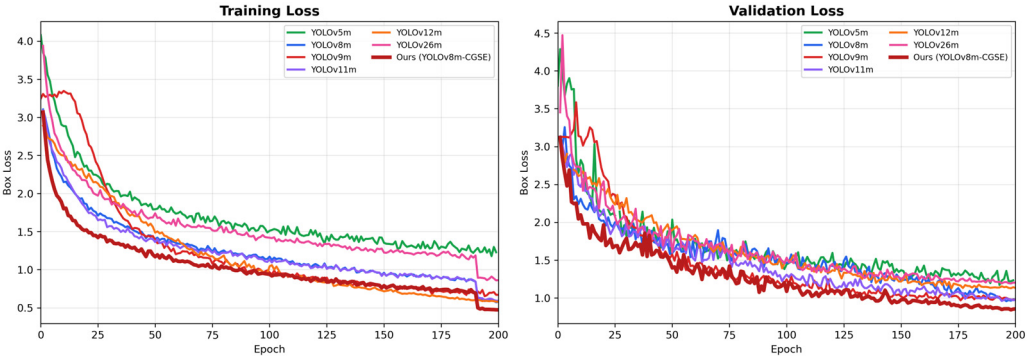


Figure 10. Comparison of loss functions between the mainstream model and our model.

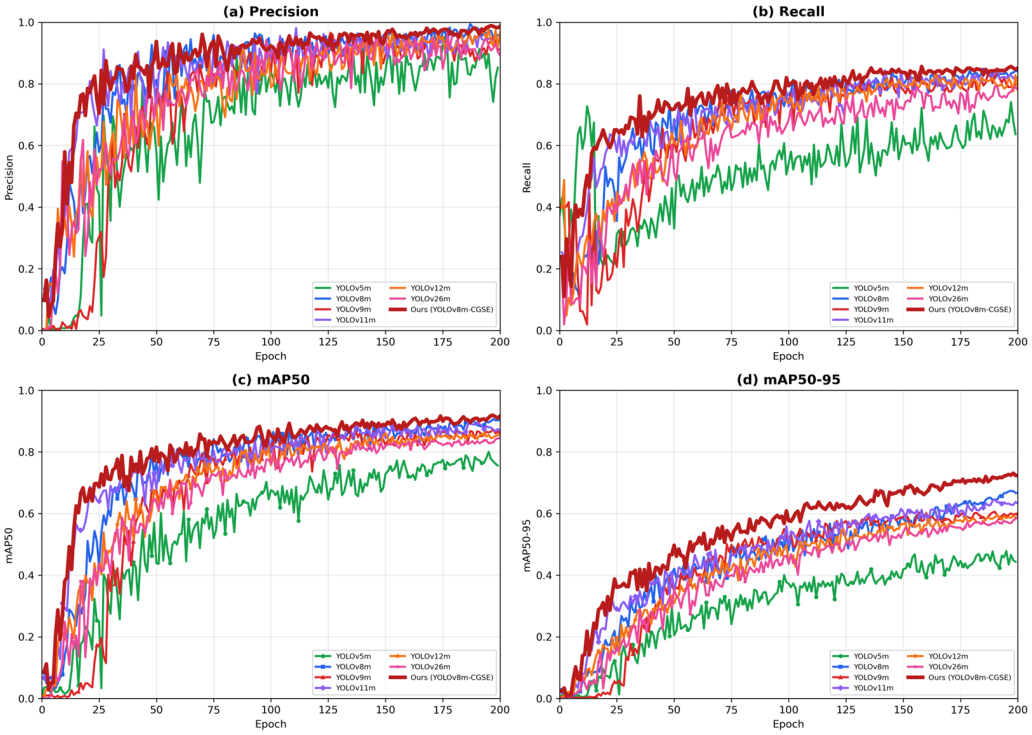


Figure 11. Comparison of accuracy between the mainstream model and our model.

Figure 10 illustrates the training and validation loss curves. It can be observed that the proposed YOLOv8m-CGSE exhibits the fastest convergence rate and reaches the lowest stable loss value compared to YOLOv5m, YOLOv8m, and even the latest YOLOv11 and YOLOv12. This indicates that the integration of GSConv and the SENetV2 attention mechanism effectively optimizes the gradient flow and enhances the model's ability to learn discriminative features from complex marine backgrounds, resulting in superior training stability.

Furthermore, Figure 11 presents the precision-recall (PR) and mAP growth curves. The results demonstrate that YOLOv8m-CGSE consistently maintains the highest precision levels throughout the training process. Particularly in scenarios involving small oil slicks and low-contrast boundaries, our model significantly outperforms the general-purpose YOLOv11 and YOLOv12 architectures. This validates that the Cross-scale Context Fusion Module (CCFM) provides a critical advantage in multi-scale feature aggregation, allowing for more robust detection in deceptive maritime environments.

YOLOv8m conducted ablation experiments by adding various modules to examine their impacts on accuracy, parameter quantity, and computational load. Eight experimental configurations were designed: (1) Baseline YOLOv8m model without any modifications; (2) Addition of the CCFM alone to validate its cross-scale feature fusion effect; (3) Replacement of traditional convolutional layers with GSConv convolutional layers to evaluate the contribution of lightweight feature extraction; (4) Addition of the SENetV2 module alone; (5) Integration of both CCFM and GSConv convolutional layers; (6) Integration of both CCFM and SENetV2 modules to explore the synergistic effect of cross-scale fusion and channel attention; (7) Integration of GSConv convolutional layers and SENetV2 modules to validate the combined value of lightweight feature extraction and attention mechanisms; (8) Integration of CCFM, GSConv convolutional layers, and SENetV2 modules (denoted as YOLOv8m-CGSE model) to assess the comprehensive performance under their synergistic effects. All experiments were conducted on the same training and validation sets, with the test set used for final evaluation. Experimental parameters remained consistent, including batch size 32,200 training rounds, and learning rate 0.001.

Table 3 summarizes the performance variations induced by different structural modules. The baseline model yields an mAP value of 0.892 with 25.84 million parameters and 78.7 G FLOPs. Incorporation of the proposed CCFM reduces both model parameters and computational cost while slightly elevating mAP50, indicating a preliminary trade-off between lightweight design and detection precision. Substitution of standard convolutions with GSConv enhances both mAP50 and mAP50-95, accompanied by reduced parameters and FLOPs, which verifies the ability of lightweight convolution in strengthening multi-scale feature representation. Introduction of SENetV2 significantly improves detection accuracy with only a minor increment in parameters, demonstrating the effectiveness of channel attention in emphasizing critical feature information. Combinations of CCFM, GSConv, and SENetV2 further promote the synergy between model compression and performance improvement. Among all configurations, the proposed YOLOv8m-CGSE model achieves the best overall performance, with notable gains in mAP50 and mAP50-95 as well as obvious reductions in parameters and FLOPs. These results validate that the integrated design effectively improves detection accuracy while maintaining high computational efficiency.

Table 3. Comparison of Ablation Experimental Results.

Name	CCFM	GSConv	SENetV2	Box (P)	R	mAP50	mAP50-95	Parameter (M)	FLOPs (G)
YOLOv8m				0.953	0.847	0.892	0.656	25.84	78.7
	√			0.971	0.847	0.895	0.656	17.14	64.4
		√		0.977	0.842	0.903	0.703	23.62	74.1
			√	0.962	0.847	0.908	0.692	25.97	78.8
	√	√		0.979	0.828	0.903	0.705	15.29	59.2
	√		√	0.977	0.847	0.903	0.689	15.29	59.2
		√	√	0.960	0.847	0.909	0.722	23.74	73.9
	√	√	√	0.981	0.847	0.912	0.733	21.69	68.8

To comprehensively evaluate the performance of the proposed YOLOv8m-CGSE, comparative experiments were extended to include not only classic detectors (YOLOv5m, SSD, Faster R-CNN) and the baseline YOLOv8m, but also the latest YOLO series architectures, namely YOLOv9m, YOLOv11m, YOLOv12m, and YOLOv26m. The quantitative comparison results are presented in Table 4. All models were trained and evaluated on the same sea surface oil spill dataset, with identical training/verification/testing set divisions, and under the hardware/software environment specified in Table 1 to ensure fairness. YOLOv5m, as the mainstream predecessor model in the YOLO series, is widely adopted for its balanced speed and accuracy. Fast R-CNN represents the classic two-stage detector, while SSD is an early representative of single-stage detectors focusing on multi-scale detection.

Table 4. Comparative Experiments of Different Models.

Name	Box (P)	R	mAP50	mAP50-95	Parameter (M)	FLOPs (G)
YOLOv5m	0.957	0.799	0.870	0.609	20.87	48.2
YOLOv8m	0.953	0.847	0.892	0.656	25.84	78.7
SSD	0.846	0.844	0.724	0.406	23.74	273.6
FasterRCNN	0.644	0.798	0.785	0.501	28.29	896.29
YOLOv9m	0.889	0.823	0.867	0.607	11.71	46.8
YOLOv11m	0.918	0.842	0.896	0.643	20.03	67.6
YOLOv12m	0.975	0.818	0.889	0.630	19.58	59.5
YOLOv26m	0.896	0.783	0.844	0.588	20.35	67.8
Ours	**0.981**	**0.847**	**0.912**	**0.733**	**21.69**	**68.8**

Table 4 presents the quantitative comparison results of various models. As shown in the table, the proposed YOLOv8m-CGSE achieves the highest detection accuracy across all core metrics, yielding a Precision of 0.981, a Recall of 0.847, an mAP50 of 0.912, and an mAP50-95 of 0.733. Compared to the classic one-stage and two-stage detectors (SSD and Faster R-CNN), all YOLO-based models demonstrate a significant advantage in balancing accuracy and computational cost. When specifically compared to the baseline YOLOv8m, our method improves mAP50 by 2.0% (from 0.892 to 0.912) and mAP50-95 by 7.7% (from 0.656 to 0.733), while concurrently reducing the parameter count by 16.1% (from 25.84 M to 21.69 M) and FLOPs by 12.6% (from 78.7 G to 68.8 G).

Furthermore, when evaluated against the newest iterations of the YOLO family (YOLOv9m, YOLOv11m, YOLOv12m, and YOLOv26m), our customized model maintains a clear competitive edge. Although recent YOLOv11m and YOLOv12m models deliver strong baseline performance (mAP@50 of 0.896 and 0.889, respectively) with efficient lightweight structures, they are typically designed for general object detection scenarios. Xu et al. [46] developed a YOLOv11-based framework for marine radar oil spill monitoring, integrated with an improved NGO algorithm to enhance the detection of weak and scattered oil features. He et al. [47] successfully deployed YOLOv12 for UAV-based marine oil spill detection and validated its ability to distinguish oil spills from complex marine backgrounds. Nevertheless, effectively detecting low-contrast and blurred-boundary oil spills in real marine environments remains a challenging task. Similarly, the YOLOv26m model achieves an mAP50 of 0.844, which falls short of our customized approach. By integrating the SENetV2 attention mechanism and the CCFM, our YOLOv8m-CGSE effectively overcomes these domain-specific challenges. The results fully validate that the proposed model achieves the most optimal balance between lightweight deployment and high-precision oil spill detection compared to both its predecessors and the latest state-of-the-art architectures.

As shown in Figure 12, a comparison of results across all columns reveals critical limitations in baseline and traditional object detection models, while demonstrating the superior robustness of YOLOv8m-CGSE in complex sea surface environments. For the slender offshore oil slick shown in Column A, models including YOLOv5m, YOLOv9m, YOLOv11m, and YOLOv26m suffer from severe missed detections, YOLOv12m exhibits highly inaccurate localization, and Faster R-CNN generates redundant bounding boxes, whereas YOLOv8m-CGSE achieves precise and complete coverage. For small nearshore oil patches in Columns B and C, SSD fails to detect small-scale oil spills and Faster R-CNN produces excessively fragmented boxes, while YOLOv8m-CGSE accurately delineates the targets without error. Under challenging illumination conditions such as strong sun glint (Column D) or diffuse gray oil films within shadowed regions (Column E), several baseline models generate loose or misaligned bounding boxes, whereas the proposed model maintains tightly fitted and well-aligned coverage. When faced with a reddish sea surface anomaly in Column F, Faster R-CNN produces false-positive detections and YOLOv26m shows minor localization deviations, yet YOLOv8m-CGSE avoids such errors entirely. For the rainbow-colored oil film in Column G, SSD, YOLOv11m, and YOLOv26m completely fail to detect the target; YOLOv5m, YOLOv8m, YOLOv9m, and YOLOv12m exhibit noticeable localization offsets; and Faster R-CNN suffers from severe positional displacement, while YOLOv8m-CGSE consistently delivers highly accurate detection and positioning. Finally, for the multi-layer iridescent film in Column H, although all models achieve roughly comparable and adequate detection performance, YOLOv8m-CGSE best maintains tightly fitting bounding boxes that closely follow the irregular edges of the oil spill.



Figure 12. (**A**–**H**) Detection of different models on different oil spill images.

Figure 13 presents the detection performance and attention mechanism visualization in typical marine oil spill scenarios, including 8 representative test scenes covering diverse oil spill characteristics and complex maritime interference factors. Row (a) shows the original input images, rows (b) and (c) display the Grad-CAM heatmaps of the baseline YOLOv8m and the proposed YOLOv8m-CGSE, respectively. It is observed that due to the irregular and fragmented shapes of marine oil films, both models tend to generate higher confidence responses along the oil film boundaries. The baseline YOLOv8m exhibits more localized and sharp attention peaks, which appear visually more concentrated but are often limited to partial regions of the oil film and accompanied by noticeable activation in non-oil regions such as sea waves and sun glint. In contrast, the YOLOv8m-CGSE model produces more continuous and spatially complete attention distributions that better align with the overall contour of the oil film and to a certain extent suppresses irrelevant background activation. This result indicates that the integration of SENetV2 and CCFM helps improve the model's ability to capture the global spatial characteristics of oil films in complex maritime backgrounds.

Overall, the visual comparison results indicate that the proposed YOLOv8m-CGSE model demonstrates superior detection accuracy, localization precision, and robustness in complex marine environments compared with conventional object detection models.

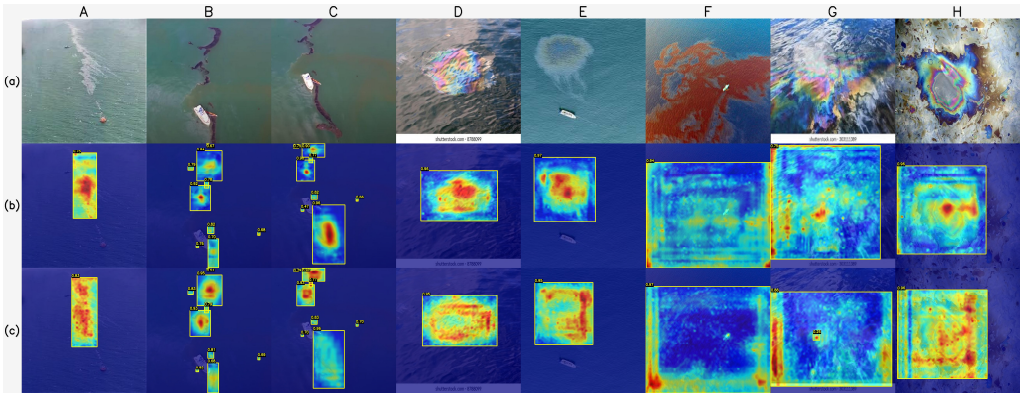


Figure 13. Heatmap comparison: (**a**) original image; (**b**) YOLOv8m prediction result; (**c**) YOLOv8m-CGSE prediction result.

4.4. Limitation Analysis of False Positives and Lookalikes

To comprehensively evaluate the boundary conditions of the proposed YOLOv8m-CGSE, particularly concerning the critical challenge of false indications (lookalikes) in visual remote sensing, a supplementary vulnerability test was conducted. As noted in authoritative remote sensing literature, optical imagery is highly susceptible to false positives caused by severe ship wakes, sun glint, and complex wave clutter, which share textural similarities with actual oil films.

As shown in Figure 14, to investigate this, we established a highly challenging negative test set comprising 30 high-resolution, oil-free marine images that exclusively feature severe ship wakes and dense wave clutter. Both the baseline YOLOv8m and the proposed YOLOv8m-CGSE were applied to this dataset to evaluate their robustness against extreme false indicators. The experimental results revealed that the baseline YOLOv8m generated false positive detections in 21 out of the 30 images (a 70% false positive rate). In contrast, the proposed YOLOv8m-CGSE reduced the false positive detections to 15 out of 30 images (a 50% false positive rate). Detailed information on dataset access is provided in the Data Availability Statement.

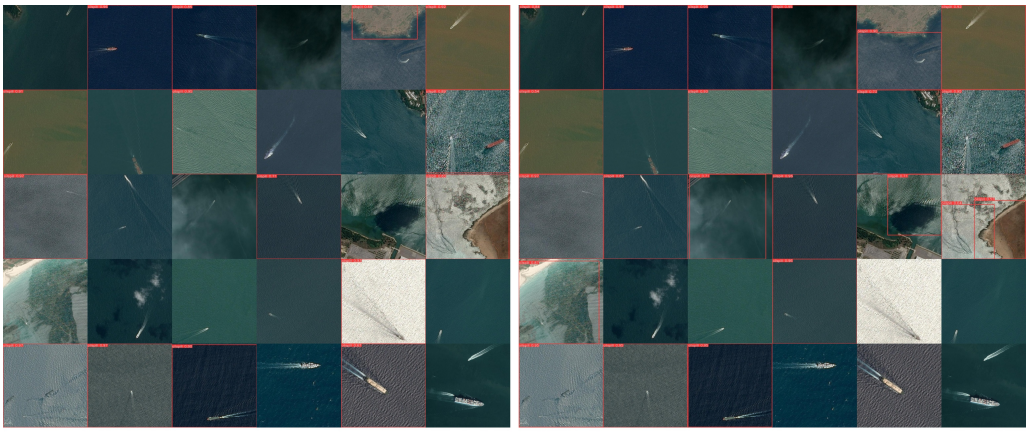


Figure 14. Oil-free marine images serve as the target for false sample detection: (**a**) YOLOv8m, (**b**) YOLOv8m-CGSE.

This comparative finding yields two important conclusions. First, the reduction in false positives demonstrates that the integration of the SENetV2 attention mechanism and the CCFM effectively enhances the model's discriminative feature extraction, actively suppressing a considerable amount of complex background clutter compared to the baseline. Second, despite this improvement, the remaining 50% false positive rate under extreme

deceptive conditions objectively highlights a primary limitation of the current study. Because our primary training dataset is strictly single-class ("oil") and focuses on distinct spill scenarios, the network implicitly treats lookalikes as background noise rather than explicitly learning to reject them through hard-negative mining.

Acknowledging this limitation is crucial for the future operational deployment of UAV optical sensors. While our model achieves state-of-the-art accuracy in typical spill scenarios, achieving zero-false-alarm monitoring in highly complex maritime environments remains a challenge. Our immediate future work will address this by constructing a comprehensive, multi-class marine dataset that explicitly incorporates diverse lookalikes as hard-negative training samples, forcing the network to learn deeper semantic discriminative features beyond simple textures.

5. Discussion

5.1. Sensor and Methodological Comparison with Existing Literature

The experimental results indicate that YOLOv8m-CGSE achieves an effective balance between detection accuracy and lightweight architecture. From a remote sensing perspective, the proposed framework shares similarities with existing deep-learning-based oil spill detection methods, as they all aim to achieve automated sea-surface anomaly identification and improve the reliability of marine environmental monitoring. However, the differences mainly lie in sensor selection, application scenarios, and model optimization strategies.

Many existing oil spill detection studies rely on satellite SAR imagery, such as the recent advanced studies reported by Shi et al. [48] and Long et al. [49]. SAR imagery has important advantages for large-scale and all-weather oil spill monitoring because it is less affected by cloud cover and illumination conditions. However, SAR-based oil spill detection may still be affected by relatively coarse spatial resolution, speckle noise, and backscatter ambiguity between true oil spills and lookalike phenomena under certain sea-state conditions. In contrast, the UAV-based optical imagery used in this study provides much higher spatial resolution and more detailed visual information, which facilitates the identification of thin oil films, fragmented slicks, and irregular spill boundaries. Therefore, our method is more suitable for local, rapid-response, and fine-scale emergency monitoring after a suspected oil spill event has been identified.

From a methodological perspective, our model is also related to recent deep learning methods for marine target detection, including UAV-based attention networks such as those proposed by He et al. [47]. Similar to these methods, YOLOv8m-CGSE improves feature representation by enhancing multi-scale feature extraction and feature fusion. However, unlike some high-complexity architectures that improve detection performance by increasing network depth, model parameters, or computational cost, the proposed model focuses on targeted lightweight optimization. Specifically, GSConv is adopted to reduce redundant convolutional computation, CCFM is introduced to enhance cross-scale contextual feature interaction, and SENetV2 is used to strengthen informative channel responses while suppressing irrelevant background interference.

With these designs, YOLOv8m-CGSE achieves an mAP50 of 91.2% and an mAP50-95 of 73.3%, while reducing the number of parameters and GFLOPs by 16.1% and 12.6%, respectively. This indicates that the proposed model does not rely on heavy model scaling to improve performance. Instead, it achieves a more practical trade-off between detection accuracy and computational efficiency. These characteristics make YOLOv8m-CGSE particularly suitable for resource-constrained UAV edge deployment and real-time marine oil spill emergency monitoring.

5.2. Robustness Against Marine Lookalikes

UAV-based optical imagery provides high spatial resolution for fine-scale marine oil spill detection, but it is also susceptible to visual interference from complex sea-surface backgrounds. In this study, particular attention was given to two common sources of interference: strong sunlight reflection and ship wakes. These phenomena may present visual patterns similar to oil films in UAV optical images, thereby increasing the risk of false positive detections.

To evaluate the robustness of the proposed model against these interferences, a targeted supplementary test was conducted using 30 deceptive oil-free images. These images contained severe visual interference caused by sunlight reflection and ship wakes, but did not include real oil spill targets. The results show that YOLOv8m-CGSE reduced the number of false positive detections from 21 in the baseline model to 15. This improvement indicates that the proposed model can more effectively suppress confusing background responses and distinguish true oil spills from visually similar interference.

Compared with models that mainly focus on improving detection accuracy for oil-containing images, this additional robustness evaluation provides a more practical assessment of false-alarm control in complex maritime environments. Although false positives were not completely eliminated, the reduction demonstrates that YOLOv8m-CGSE improves operational reliability under challenging UAV-based marine monitoring conditions.

5.3. Limitations and Future Work

Despite current improvements, the model's detection performance in extremely harsh marine environments—characterized by low target contrast and severe interference—requires further enhancement. Furthermore, because the dataset focuses exclusively on the "oil" category, the network implicitly treats marine "lookalikes" as background noise rather than explicitly classifying them, which may still lead to false positives in highly deceptive scenarios. To address these limitations and improve geographical generalization, future work will focus on constructing a comprehensive, multi-class dataset that explicitly includes various lookalike categories. This will enable the model to directly learn discriminative features between true oil spills and environmental artifacts. Additionally, we will continue optimizing the lightweight architecture to facilitate seamless deployment in real-time, UAV-based monitoring systems.

6. Conclusions

This study proposed YOLOv8m-CGSE, a lightweight marine oil spill detection model designed for UAV-based monitoring under complex background interference and resource-constrained deployment conditions. By integrating GSConv, CCFM, and SENetV2 into the YOLOv8m framework, the proposed architecture enhances multi-scale feature representation, strengthens cross-scale contextual information interaction, and improves the discrimination of oil spill targets from complex marine backgrounds.

Quantitatively, YOLOv8m-CGSE achieves an mAP50 of 91.2% and an mAP50-95 of 73.3%, representing improvements of 2.0% and 7.7% over the baseline model, respectively. Meanwhile, the model reduces the number of parameters by 16.1% and computational cost by 12.6%, demonstrating that improved detection performance can be achieved without increasing model complexity. In the robustness test using deceptive oil-free images, YOLOv8m-CGSE reduced false positive detections from 21 to 15, indicating improved background suppression capability against marine lookalikes.

Compared with many SAR-based oil spill detection methods, the proposed UAV optical framework provides finer spatial details and is more suitable for local, rapid-response, and fine-scale monitoring. Compared with some computationally intensive

optical detection models, YOLOv8m-CGSE achieves a better balance between accuracy, model complexity, and deployment feasibility. Therefore, the main benefits of this work are threefold: first, it improves the detection accuracy of marine oil spills in UAV imagery; second, it reduces model parameters and computational cost, making it more suitable for edge deployment; and third, it enhances robustness against visually similar marine interference, thereby reducing false alarms.

Overall, the proposed YOLOv8m-CGSE provides a high-precision, low-false-alarm, and computationally efficient solution for real-time UAV-based marine oil spill emergency monitoring. Future work will focus on constructing a multi-class dataset containing both true oil spills and typical marine lookalikes and validating the model on real UAV edge-computing platforms to further improve its practical applicability.

Author Contributions: Conceptualization, Q.W. and J.L. (Junjie Lu); methodology, Q.W. and J.L. (Junjie Lu); validation, Q.W., J.L. (Junjie Lu) and B.Y.; formal analysis, J.L. (Junjie Lu), J.L. (Jingwen Li) and T.Y.; investigation, Q.W., B.S. and J.L. (Junjie Lu); resources, C.J., J.J. and B.Y.; data curation, Q.W. and B.Y.; writing—original draft preparation, Q.W.; writing—review and editing, J.L. (Junjie Lu), G.Z. and J.L. (Jingwen Li); funding acquisition, Q.W. and G.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Young and Middle-aged Teachers' Basic Scientific Research Ability Improvement Project (No. 2024KY0281), Qingmiao Talent Research Project Funding (Qingyang Wang), National Natural Science Foundation of China (Project No. 42461050), Guangxi Key Laboratory of Spatial Information and Geomatics Program (No. 21-238-21-24), and Guilin University of Technology (No. RD2300151852).

Data Availability Statement: The marine oil spill data used in this study are available at: https://universe.roboflow.com/baka-toast/oilspill-one/dataset/1 (accessed on 27 October 2025). The boat wake image data used in this study are available at: https://universe.roboflow.com/object-detection-yno11/boat-wake-detection1-ponpo/dataset/1 (accessed on 10 May 2026).

Acknowledgments: We thank all authors for their support in conducting this research.

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

References

1. Michel, J.; Fingas, M. Oil spills: Causes, consequences, prevention, and countermeasures. In *Fossil Fuels: Current Status and Future Directions*; World Scientific Publishing Company: Singapore, 2016; pp. 159–201.
2. Zhang, Z.; Sun, H.; Guo, Y. The Impact of Marine Oil Spills on the Ecosystem. *Int. J. Eng. Sci. Technol.* **2024**, *2*, 1–10. [CrossRef]
3. Fingas, M.F.; Brown, C.E. Review of Oil Spill Remote Sensing. *Spill Sci. Technol. Bull.* **1997**, *4*, 199–208. [CrossRef]
4. Burgués, J.; Marco, S. Environmental chemical sensing using small drones: A review. *Sci. Total Environ.* **2020**, *748*, 141172. [CrossRef]
5. Xu, L.; Li, J.; Brenning, A. A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. *Remote Sens. Environ.* **2013**, *141*, 14–23. [CrossRef]
6. Mera, D.; Bolon-Canedo, V.; Cotos, J.; Alonso-Betanzos, A. On the use of feature selection to improve the detection of sea oil spills in SAR images. *Comput. Geosci.* **2017**, *100*, 166–178. [CrossRef]
7. Konik, M.; Bradtke, K. Object-oriented approach to oil spill detection using ENVISAT ASAR images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *118*, 37–52. [CrossRef]
8. Topouzelis, K.; Psyllos, A. Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 135–143. [CrossRef]
9. Tong, S.; Liu, X.; Chen, Q.; Zhang, Z.; Xie, G. Multi-feature based ocean oil spill detection for polarimetric SAR data using random forest and the self-similarity parameter. *Remote Sens.* **2019**, *11*, 451. [CrossRef]
10. Park, S.; Jung, H.; Lee, M. Oil spill mapping from Kompsat-2 high-resolution image using directional median filtering and artificial neural network. *Remote Sens.* **2020**, *12*, 253. [CrossRef]
11. Topouzelis, K.; Karathanassi, V.; Pavlakis, P.; Rokos, D. Detection and discrimination between oil spills and look-alike phenomena through neural networks. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 264–270. [CrossRef]

12. Singha, S.; Bellerby, T.J.; Trieschmann, O. Satellite oil spill detection using artificial neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2355–2363. [CrossRef]
13. Bianchi, F.M.; Espeseth, M.M.; Borch, N. Large-scale detection and categorization of oil spills from SAR images with deep learning. *Remote Sens.* **2020**, *12*, 2260. [CrossRef]
14. Zhao, S.; Zhou, H.; Yang, H. Smart monitoring method for land-based sources of marine outfalls based on an improved YOLOv8 model. *Water* **2024**, *16*, 3285. [CrossRef]
15. Chai, Y.; Han, X.; Wang, Y.; Luo, D.; Yang, J.; Chen, P.; Zheng, G. *TransOilSeg*: A novel SAR oil spill detection method addressing data limitations and look-alike confusions. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5206216. [CrossRef]
16. Li, J.; Ma, Y.; Ji, Y.; Jiang, Z.; Du, K.; Liu, R.; Yang, J. SR-SqueezeNet: A lightweight hyperspectral identification model for oil spills based on smoothed activation functions. *Mar. Pollut. Bull.* **2025**, *211*, 117365. [CrossRef] [PubMed]
17. Wang, J.; Zhao, H. Improved YOLOv8 algorithm for water surface object detection. *Sensors* **2024**, *24*, 5059. [CrossRef]
18. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
19. Trigka, M.; Dritsas, E. A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection. *Sensors* **2025**, *25*, 214. [CrossRef]
20. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef] [PubMed]
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.
22. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
23. Nawarathne, U.; Kumari, H.; Kumari, H. Underwater waste detection using deep learning: A performance comparison of YOLOv7 to v10 and Faster R-CNN. *arXiv* **2025**, arXiv:2507.18967.
24. Gao, Y.; Wu, C.; Ren, M.; Feng, Y. Refined anchor-free model with feature enhancement mechanism for ship detection in infrared images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 12946–12960. [CrossRef]
25. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [CrossRef]
26. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856. [CrossRef]
27. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [CrossRef]
28. Li, M.; Wang, J.; Chen, S.; Liu, L.; Li, K.; Zhao, Z.; Yun, H. A structurally optimized and efficient lightweight object detection model for autonomous driving. *Sensors* **2025**, *26*, 54. [CrossRef] [PubMed]
29. Zhang, W.; Yi, D.; Fang, Z.; Zhao, Y.; You, Z. Research on complex defect detection method on steel surface based on EBA-YOLO. *IAENG Int. J. Comput. Sci.* **2025**, *52*, 2141–2151.
30. Chen, Z.; Lu, S. CAF-YOLO: A robust framework for multi-scale lesion detection in biomedical imagery. In Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing, Hyderabad, India, 6–11 April 2025; pp. 1–5.
31. Yu, H.; Luo, Q.; Peng, W.; Zheng, L.; Ju, J.; Zhuo, H. PKD-YOLOv8: A collaborative pruning and knowledge distillation framework for lightweight rapeseed pest detection. *Sensors* **2025**, *25*, 5004. [CrossRef]
32. Zhu, J.; Hu, T.; Zheng, L.; Zhou, N.; Ge, H.; Hong, Z. YOLOv8-C2f-Faster-EMA: An improved underwater trash detection model based on YOLOv8. *Sensors* **2024**, *24*, 2483. [CrossRef]
33. Rao, W.; Hu, Q.; Chen, G. Research on a lightweight algorithm for seabed organism detection based on deep learning. *J. Mar. Sci. Eng.* **2026**, *14*, 454. [CrossRef]
34. Sun, H.; Zhao, H.; Liu, Z.; Jiang, G.; Zhao, J. WA-YOLO: Water-aware improvements for maritime small-object detection under glare and low-light. *J. Mar. Sci. Eng.* **2025**, *14*, 37. [CrossRef]
35. Brekke, C.; Solberg, A.H.S. Oil Spill Detection by Satellite Remote Sensing. *Remote Sens. Environ.* **2005**, *95*, 1–13. [CrossRef]
36. Leifer, I.; Lehr, W.J.; Simecek-Beatty, D.; Bradley, E.; Clark, R.; Dennison, P.; Hu, Y.; Matheson, S.; Jones, C.E.; Holt, B.; et al. State of the Art Satellite and Airborne Marine Oil Spill Remote Sensing: Application to the BP Deepwater Horizon Oil Spill. *Remote Sens. Environ.* **2012**, *124*, 185–209. [CrossRef]
37. Topouzelis, K.N. Oil Spill Detection by SAR Images: Dark Formation Detection, Feature Extraction and Classification Algorithms. *Sensors* **2008**, *8*, 6642–6659. [CrossRef]
38. Temitope Yekeen, S.; Balogun, A.-L. Advances in Remote Sensing Technology, Machine Learning and Deep Learning for Marine Oil Spill Detection, Prediction and Vulnerability Assessment. *Remote Sens.* **2020**, *12*, 3416. [CrossRef]

39. Cai, Y.; Chen, L.; Zhuang, X.; Zhang, B. Automated marine oil spill detection algorithm based on single-image generative adversarial network and YOLOv8 under small samples. *Mar. Pollut. Bull.* **2024**, *203*, 116475. [CrossRef]
40. Wang, Y.; Yang, H.; Gao, L.; Wang, M.; Wang, L. Research on target detection algorithm for sea surface oil spill recognition in SAR image on YOLOS—Taking improved YOLOv8 as an example. *IAENG Int. J. Comput. Sci.* **2025**, *52*, 4952–4962.
41. Aggarwal, P.; Gangwar, P.; Verma, T.; Jindal, S.; Mohapatra, A.K.; Gupta, A. Experimental evaluation and validation of deep learning-based approach for oil spill detection in sea. *Int. J. Remote Sens.* **2025**, *46*, 7639–7655. [CrossRef]
42. Sudani, A.A.I.; Suhail, A.A.G. ResNet-OSD: An optimized hybrid deep learning framework for oil spill detection in coastal drone imagery. *Vis. Comput.* **2026**, *42*, 165. [CrossRef]
43. Dong, X.; Li, J.; Li, B.; Jin, Y.; Miao, S. Marine oil spill detection from low-quality SAR remote sensing images. *J. Mar. Sci. Eng.* **2023**, *11*, 1552. [CrossRef]
44. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
45. Yekeen, S.T.; Balogun, A.L. Automated marine oil spill detection using deep learning instance segmentation model. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 1271–1276. [CrossRef]
46. Xu, J.; Huang, Y.; Yan, J.; Guo, Z.; Li, B.; Dong, H.; Liu, P. Marine Radar Oil Spill Monitoring Method Based on YOLOv11 and Improved NGO Algorithm. *Remote Sens.* **2025**, *17*, 3922. [CrossRef]
47. He, L.; Zhou, Y.; Yang, H.; Su, L.; Ma, J. A Deep Learning-Based Method for Marine Oil Spill Detection and Its Application in UAV Imagery. *Mar. Pollut. Bull.* **2026**, *222*, 118889. [CrossRef] [PubMed]
48. Shi, J.; Jiao, T.; Ames, D.P.; Chen, Y.; Xie, Z. Improved Lightweight Marine Oil Spill Detection Using the YOLOv8 Algorithm. *Appl. Sci.* **2026**, *16*, 780. [CrossRef]
49. Cai, Y.; Su, J.; Song, J.; Xu, D.; Zhang, L.; Shen, G. LRA-UNet: A Lightweight Residual Attention Network for SAR Marine Oil Spill Detection. *J. Mar. Sci. Eng.* **2025**, *13*, 1161. [CrossRef]

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