

Article

SDE-Net: A Strip-Directional Dynamic-Scale and Edge-Aware Network for SAR Oil Spill Segmentation

Yifei Shen ¹, Yijing Liu ² , Guoru Li ³, Shuxi Chen ¹, Yuanzhi Zhang ^{4,5} and Shentao Wang ^{1,*} ¹ School of Artificial Intelligence, Nantong Institute of Technology, Nantong 226001, China; 2310515035@ntit.edu.cn (Y.S.)² School of Yonyou Digital & Intelligence, Nantong Institute of Technology, Nantong 226001, China³ School of Software, Northwestern Polytechnical University, Xi'an 710129, China⁴ School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China⁵ Faculty of Social Science, Chinese University of Hong Kong, Hong Kong 999077, China

* Correspondence: wangshentao@ntit.edu.cn

Abstract

Marine oil spill monitoring plays a critical role in environmental protection and emergency response. Synthetic aperture radar (SAR) provides all-day, all-weather, and wide-area imaging capabilities and has become a primary sensor for operational oil spill surveillance. Nevertheless, SAR oil spill segmentation remains challenging. Real slicks are easily confused with look-alike dark formations; oil film boundaries are weak and fragmented under speckle noise; and small and sparse targets, such as ships, are easily overlooked amid the dominant sea surface background and complex coastal structures. To address these challenges, this paper proposes SDE-Net, an encoder–decoder network built on a ConvNeXtV2-Tiny backbone and a UPerNet-style decoder with three task-oriented modules. The Strip-Directional Local Enhancement (SDLE) module refines elongated low-contrast cues in the shallow lateral features of C_2 and C_3 while limiting the indiscriminate enhancement of speckle-contaminated responses. The Dynamic Scale Pyramid Context (DSPC) module redesigns pyramid-based context aggregation at the deepest stage by introducing an input-conditioned softmax scale gate that adaptively reweights the contributions of predefined contextual branches. The Edge-Aware Dynamic Gated FPN (EDG-FPN) decoder combines DySample-based feature alignment and dynamic gated fusion with a boundary-aware gate driven by an edge prediction branch, modulating high-level semantic propagation in boundary-sensitive regions without requiring additional edge annotations. On the Oil Spill Detection Dataset, SDE-Net achieves a 72.94% mIoU and 82.88% mDice, outperforming the UPerNet baseline equipped with the same ConvNeXtV2-Tiny backbone by 4.25 and 3.84 percentage points, respectively. The results demonstrate improved oil spill delineation and the preservation of sparse small-class structures under complex SAR sea surface conditions.



Academic Editor: Merv Fingas

Received: 25 July 2026

Revised: 10 September 2026

Accepted: 14 September 2026

Published: 15 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.**Keywords:** SAR image segmentation; marine oil spill; ConvNeXtV2

1. Introduction

Marine oil spills are abrupt environmental accidents caused by offshore drilling, tanker collisions, pipeline leakage, and illegal discharges. Once released, oil slicks can spread rapidly under the combined forcing of wind, waves, and ocean currents, causing persistent damage to marine ecosystems, fishery resources, port operations, coastal industries, and human livelihoods [1–6]. Effective emergency response therefore depends on the timely

acquisition of the location, extent, boundary morphology, and drift tendency of the oil slick. Although vessel patrols, buoy observations, and airborne surveys can provide accurate local measurements, their high costs, weather sensitivity, and limited spatial coverage restrict their scalability in large-area emergency monitoring [7–9]. Synthetic aperture radar (SAR), by contrast, provides all-day, all-weather, and wide-swath imaging capabilities, making it one of the most important remote sensing tools for operational marine oil spill surveillance [10,11]. Oil films dampen capillary and short gravity waves on the sea surface, thereby reducing radar backscatter and forming dark regions in SAR imagery [12–14]. This imaging mechanism enables SAR to detect suspected oil spill areas even under cloudy, rainy, or nighttime conditions.

For oil spill emergency interpretation, pixel-level semantic segmentation is more informative than image-level classification or bounding box detection because it directly delineates the fine contours, spatial extent, area, and morphological pattern of the slick [15,16]. Two-stage deep learning frameworks combine SAR oil spill detection and segmentation [17], while the Coastal Crete digital twin integrates these tasks with environmental forcing and Lagrangian drift forecasting to support emergency response [18]. More accurate oil masks and boundaries can better represent the observed slick extent used to initialize such forecasts, providing application rationale for improving the intersection over union for the oil spill class (Oil IoU) and boundary delineation. These outputs also support accident severity estimation, response scheduling, pollution containment, and damage assessment [16,19]. Early SAR oil spill studies mainly relied on adaptive thresholding, region growing, level sets, and handcrafted geometric or textural descriptors, followed by statistical classifiers such as support vector machines, random forest, and Mahalanobis distance classifiers [10,20,21]. These methods are interpretable and require limited labeled data, but their performance is strongly dependent on the sea state, the imaging geometry, speckle noise, and manually designed feature representations [22,23]. As a result, they usually perform reliably only in regular sea conditions and high-contrast scenarios [23,24].

The main obstacle in SAR oil spill segmentation is not merely the extraction of dark regions but reliable discrimination between real oil spills and visually similar look-alike phenomena [8,21]. Low-wind areas, biogenic films, oceanic eddies, algal blooms, ship wakes, and organic surface films can produce dark signatures that resemble those of oil slicks in SAR imagery [8,15,21]. Such ambiguity makes fixed thresholds and shallow shape descriptors insufficient under variable maritime conditions, often resulting in false alarms, missed detections, and poor cross-scene generalization [8,21]. The launch of Sentinel-1 in 2014, together with the release of public oil spill datasets, has promoted a transition from handcrafted feature-based pipelines to data-driven segmentation methods [15,21,25]. In particular, the Oil Spill Detection Dataset (OSDD), constructed from Sentinel-1 imagery and EMSA CleanSeaNet pollution records, provides pixel-level annotations for five maritime categories—sea surface, oil spill, look-alike, ship, and land [15]. Owing to its public availability and fine-grained category annotations, the OSDD has become a commonly used benchmark for evaluating SAR oil spill segmentation methods [26,27].

Deep learning has therefore become the dominant direction for SAR oil spill segmentation. Fully convolutional networks and U-Net established the basic encoder–decoder paradigm for dense prediction [28,29], while subsequent methods focused on improving structural stability, scale robustness, and boundary localization. Sequence-based dark spot recognition, OSCNet, DeepLabV3+-based contextual aggregation, Bayesian optimization strategies, and few-shot generative models such as MTGAN have been investigated to improve the segmentation accuracy under limited labeled samples and complex scale variations [30–34]. More recent studies have further enhanced feature discrimination through multiconvolutional context modules, polarization-aware attention, dual atten-

tion mechanisms, and boundary-enhanced designs [35–38]. FCS-Net, developed in an earlier work involving one of the present authors, combines frequency–spatial attention with strip-augmented context aggregation [39]. Whereas its strip pooling operates at the deepest bottleneck, SDE-Net applies directional refinement to shallow-to-mid-level lateral features and coordinates it with input-adaptive deep context aggregation and boundary-sensitive cross-level fusion. These studies indicate that SAR oil spill segmentation requires not only stronger feature extraction but also more effective context modeling and class-sensitive representation.

Meanwhile, attention-based and Transformer-based architectures have shown strong potential in remote sensing and maritime interpretation tasks. Hybrid Transformer models such as OSDTAU-Net exploit global attention to capture long-range dependencies that are difficult for pure convolutional networks to model [40]. Foundation model-assisted methods include SAM-OIL and a recent two-stage framework combining ConvNeXt-T screening with domain-adapted OSDA-SAM segmentation [41,42]. SDE-Net complements this direction with a task-specific decoder trained using pixel-wise supervision and mask-derived boundary targets, without additional edge annotations. In addition, class imbalance remains a persistent issue because oil spill pixels often occupy only a small proportion in SAR images. To address this problem, two-stage preprocessing strategies, filtering operations, and imbalance-aware losses such as the generalized Dice loss have been introduced [43]. Lightweight models such as FA-MobileUNet have also been explored to reduce the operational cost while maintaining segmentation capabilities [44]. These developments suggest that modern SAR oil spill segmentation has gradually converged toward three design pillars: a strong hierarchical encoder, an effective multiscale context aggregator, and a discriminative cross-level fusion decoder.

Despite this progress, several key challenges remain unresolved in real SAR sea surface scenes. First, real oil films and non-oil dark formations share similar low-backscatter responses, so look-alike confusion remains a persistent source of false alarms, with the most ambiguous cases concentrating along class transitions where local appearance cues are nearly identical [45]. Second, oil slicks usually exhibit anisotropic and filamentary structures with thin, curved, and fragmented boundaries, which are difficult to capture with standard isotropic convolutional kernels [39]. Third, the boundary contrast between oil films and the surrounding sea surface is often weak and unstable due to speckle noise, incidence angle variations, and heterogeneous sea states, further amplifying both look-alike confusion and boundary localization errors [38,45]. Fourth, the interaction between directional shallow cues, deep contextual representations, and boundary-sensitive semantic propagation requires coordinated design across feature levels. Therefore, a task-oriented segmentation framework is required to integrate shallow structural refinement, input-adaptive context aggregation, and boundary-aware cross-level fusion in a unified manner.

To address these difficulties, this paper builds upon a ConvNeXtV2–UPerNet segmentation framework and develops three task-oriented enhancements that target the above challenges in a complementary manner. ConvNeXtV2-Tiny is adopted as the hierarchical encoder to extract multistage representations from SAR imagery, while the UPerNet-style decoding pipeline is redesigned to better accommodate the anisotropic morphology, weak boundary contrast, and look-alike ambiguity of oil slicks [46,47]. First, the Strip-Directional Local Enhancement (SDLE) module strengthens elongated, low-contrast structures in lateral features derived from C_2 and C_3 . Second, the Dynamic Scale Pyramid Context (DSPC) module adjusts the contributions of pooled features at different scales according to each input image. Third, the Edge-Aware Dynamic Gated FPN (EDG-FPN) decoder combines features from different levels using weights predicted from their content. EDG-FPN uses DySample [48] for feature alignment and a boundary-aware gate (BAG) to guide fusion at

the highest-resolution stage. The gate uses predicted boundaries learned from targets generated directly from segmentation labels [49,50]. Additional edge annotations are therefore unnecessary for training this boundary prediction branch.

The main contributions of this paper are summarized as follows:

1. We design a Strip-Directional Local Enhancement (SDLE) module for shallow-level feature refinement in SAR oil spill segmentation. By coupling local, horizontal, and vertical depth-wise operations with a spatial gate and a learnable near-identity residual, SDLE strengthens elongated low-contrast oil spill cues while preserving the stability of the original lateral features.
2. We introduce a Dynamic Scale Pyramid Context (DSPC) module for input-adaptive context aggregation at the deepest stage of UPerNet. For each input image, a soft-max gate assigns weights to the predefined pooling branches, allowing their relative contributions to vary across scenes.
3. We develop an Edge-Aware Dynamic Gated FPN (EDG-FPN) decoder for content-adaptive and boundary-sensitive cross-level fusion. EDG-FPN combines DySample-based feature alignment with dynamic gated propagation and a boundary-aware gate driven by an edge prediction branch trained using on-the-fly boundary targets derived from segmentation labels. This design adaptively regulates high-level semantic propagation in boundary-sensitive regions without requiring additional edge annotations.
4. Extensive experiments on the OSDD demonstrate that SDE-Net achieves a 72.94% mIoU and 82.88% mDice, outperforming the UPerNet baseline equipped with the same ConvNeXtV2-Tiny backbone by 4.25 and 3.84 percentage points, respectively.

2. Methodology

2.1. Overall Architecture

As illustrated in Figure 1, the proposed SDE-Net follows a pyramid-based encoder-decoder segmentation framework [47] built on a ConvNeXtV2-Tiny backbone. The backbone blocks use global response normalization (GRN) to enhance interchannel feature competition [46]. Given an input SAR image $X \in \mathbb{R}^{3 \times H \times W}$, the encoder extracts four hierarchical feature maps $\{C_1, C_2, C_3, C_4\}$ with strides $\{4, 8, 16, 32\}$ and channel dimensions $\{96, 192, 384, 768\}$, respectively [46]. Among them, C_1 retains fine spatial details and local boundary responses, C_2 and C_3 provide shallow-to-mid-level structural representations, and C_4 carries the strongest semantic context with the largest receptive field. These multistage features form the input to the subsequent task-oriented decoding pathway.

The decoding pathway is redesigned at three key positions. After the lateral 1×1 projections that unify C_1 , C_2 , and C_3 to a common decoder width $C_d = 256$, the lateral features of C_2 and C_3 are refined by the Strip-Directional Local Enhancement (SDLE) module, while the lateral feature of C_1 is passed through without SDLE refinement to avoid overamplifying noise-sensitive low-level responses. The 768-channel feature C_4 is processed directly by the Dynamic Scale Pyramid Context (DSPC) module to produce a 256-channel output P_4 through input-adaptive context aggregation before top-down decoding. Finally, the resulting multilevel features are fused by the Edge-Aware Dynamic Gated FPN (EDG-FPN) decoder, which combines DySample-based feature alignment [48] with dynamic gated top-down fusion at the deeper stages and a boundary-aware gate at the highest-resolution stage. The boundary-aware gate is modulated by an edge prediction branch supervised using on-the-fly boundary targets derived from segmentation labels [51], thereby adaptively regulating high-level semantic propagation in boundary-sensitive regions without requiring additional edge annotations.

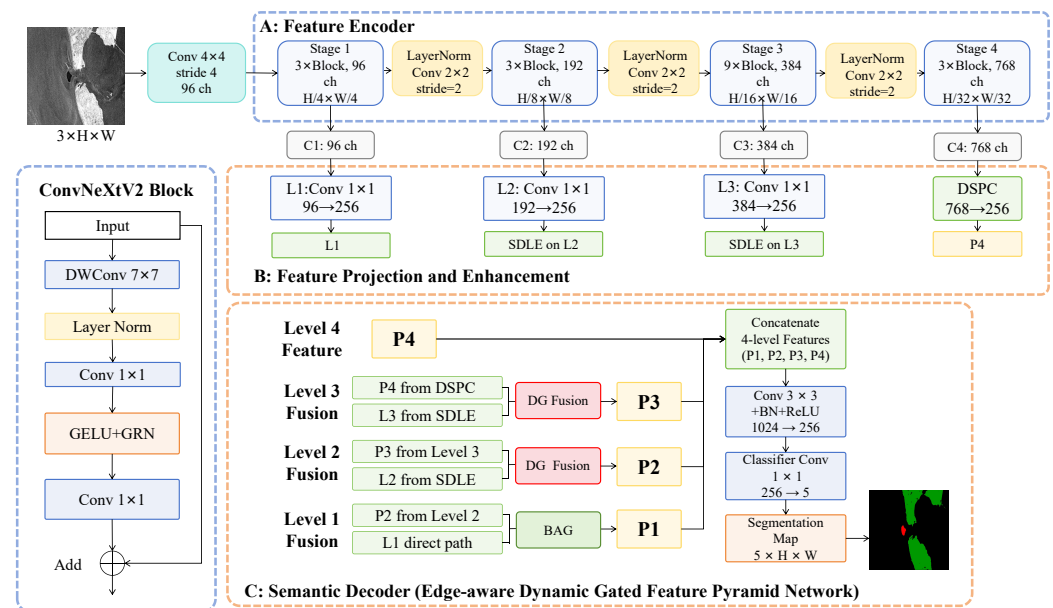


Figure 1. Overall architecture of the proposed SDE-Net. The ConvNeXtV2-Tiny encoder extracts four hierarchical features, where C_2 and C_3 are refined by SDLE and C_4 is enhanced by DSPC. The resulting multilevel features are fused by the semantic decoder to generate the final five-class segmentation map.

After multilevel fusion, the aggregated feature is passed through a segmentation classifier to produce dense logits $Y \in \mathbb{R}^{K \times H \times W}$, where $K = 5$ corresponds to the five semantic categories in the Oil Spill Detection Dataset: sea surface, oil spill, look-alike, ship, and land [15]. The final prediction is obtained by assigning each pixel to the class with the highest logit response. Within this architecture, SDLE enhances elongated local structures at shallow levels, DSPC adapts the relative contributions of high-level contextual scales to each input scene, and EDG-FPN integrates spatial details with semantic context in a boundary-aware manner.

2.2. Strip-Directional Local Enhancement Module

Oil slicks in SAR imagery often appear as elongated, low-contrast, and locally fragmented dark structures. Their boundaries are usually thin and unstable because of speckle noise, heterogeneous sea states, and local backscatter fluctuations. In a pyramid-based encoder–decoder segmentation framework, shallow features preserve fine spatial details but may also contain strongly noise-sensitive responses. Consequently, applying strip enhancement indiscriminately to all low-level features may introduce unnecessary interference. To refine elongated oil slick cues while limiting the amplification of noisy responses, we introduce the Strip-Directional Local Enhancement (SDLE) module for selected shallow-to-mid-level lateral features, as illustrated in Figure 2.

In the proposed architecture, SDLE is applied only to the lateral features derived from C_2 and C_3 , rather than to all encoder stages. The lowest-level feature C_1 has the highest spatial resolution and contains abundant local edge and texture responses, but it is also the most sensitive to speckle noise and unstable local scattering patterns. Applying strip enhancement directly to C_1 may amplify noisy dark fragments and increase the risk of confusing oil slicks with visually similar dark regions. By contrast, C_2 and C_3 retain sufficient structural information while providing more stable semantic cues than C_1 . Accordingly, SDLE is applied to the C_2 - and C_3 -derived lateral features, whereas the C_1 lateral feature is retained as a direct path to the decoder.

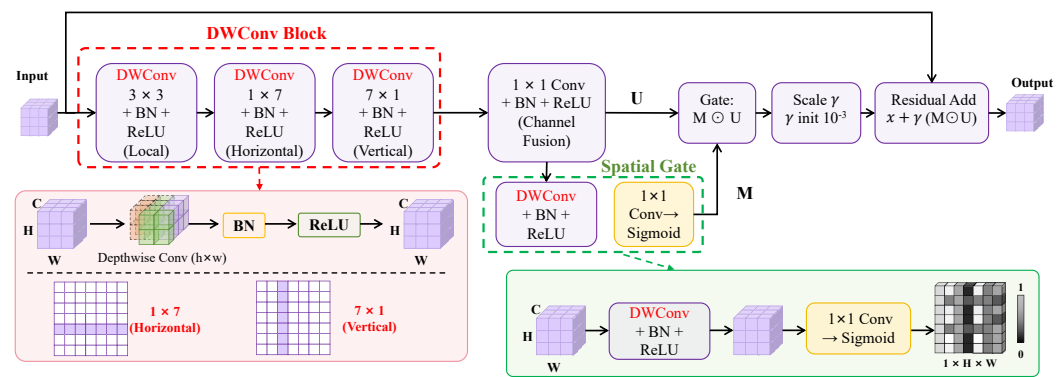


Figure 2. Structure of the proposed Strip-Directional Local Enhancement (SDLE) module. Given an input lateral feature x , SDLE first applies a local 3×3 depth-wise convolution, followed by horizontal $1 \times k$ and vertical $k \times 1$ depth-wise convolutions for axis-wise strip aggregation, where $k = 7$. The aggregated response is projected by a 1×1 convolution to obtain U . A single-channel spatial gate M is then predicted from U and broadcast along the channel dimension to produce the gated response $M \odot U$. Finally, the enhanced response is injected into the input through a learnable residual scale γ , initialized to 10^{-3} , yielding $x + \gamma(M \odot U)$.

Given a channel-aligned lateral feature $x \in \mathbb{R}^{B \times C_d \times H \times W}$, SDLE first extracts local responses and then performs sequential axis-wise strip aggregation:

$$\begin{aligned} u_1 &= \phi(\text{DWConv}_{3 \times 3}(x)), \\ u_2 &= \phi(\text{DWConv}_{1 \times k}(u_1)), \\ u_3 &= \phi(\text{DWConv}_{k \times 1}(u_2)), \end{aligned} \quad (1)$$

where $k = 7$ in our implementation and $\phi(\cdot) = \text{ReLU}(\text{BN}(\cdot))$. The initial 3×3 depth-wise convolution extracts short-range local patterns, while the subsequent $1 \times k$ and $k \times 1$ depth-wise convolutions progressively aggregate the responses along the horizontal and vertical axes. All convolutions operate with a stride of one and size-preserving padding, so the spatial resolution remains unchanged throughout the module.

The directionally aggregated response is fused using a point-wise projection:

$$U = \phi(\text{Conv}_{1 \times 1}(u_3)), \quad U \in \mathbb{R}^{B \times C_d \times H \times W}. \quad (2)$$

A single-channel spatial gate is then predicted from U :

$$M = \sigma(\text{Conv}_{1 \times 1}(\phi(\text{DWConv}_{3 \times 3}(U)))), \quad M \in [0, 1]^{B \times 1 \times H \times W}, \quad (3)$$

where $\sigma(\cdot)$ denotes the sigmoid function, and the final 1×1 convolution maps the C_d -channel response to a single-channel spatial mask. The gate M is broadcast along the channel dimension and multiplied element-wise with U to suppress spatially irrelevant responses while retaining locations that exhibit useful directional structure.

The final output is defined as a conservative residual refinement:

$$\text{SDLE}(x) = x + \gamma(M \odot U), \quad (4)$$

where $\odot$ denotes element-wise multiplication and γ is a learnable residual scale initialized to 10^{-3} . The small initial value of γ makes the initial output close to x , while allowing the strength of the added response to be learned during training.

Although SDLE employs horizontal and vertical strip convolutions, it is not restricted to strictly horizontal or vertical oil slick structures. In a discrete SAR image, an oblique or curved elongated target can be represented by a sequence of locally connected responses

distributed across both image axes. The initial local operation captures short-range structural cues, while the subsequent horizontal and vertical strip operations accumulate these cues along complementary coordinate directions. Consequently, oblique and mildly curved elongated structures can also benefit from their accumulated horizontal and vertical projections. SDLE should therefore be understood as a lightweight anisotropic refinement module rather than a strict horizontal–vertical detector.

SDLE differs from the Strip-Augmented Pyramid Pooling (SAPP) module of FCS-Net [39] in both its aggregation operation and its position in the network. SAPP combines full-axis horizontal and vertical strip pooling with pyramid pooling at the deepest bottleneck to aggregate long-range context. SDLE instead applies sequential 3×3 , 1×7 , and 7×1 depth-wise convolutions to the C_2 - and C_3 -derived lateral features, together with spatial gating and a near-identity residual update. This design involves local directional refinement before pyramid fusion while preserving the spatial resolution of the lateral features. Deep contextual aggregation is handled separately by DSPC.

2.3. Dynamic Scale Pyramid Context Module

High-level contextual information is essential for SAR oil spill segmentation because oil slicks often appear under complex sea surface conditions and exhibit substantial variations in spatial extent. Some oil films appear as compact dark patches, some are fragmented into narrow streaks, and others extend over large irregular regions. Moreover, look-alike formations may share similar low-backscatter signatures with real oil slicks, making the local appearance alone insufficient for reliable interpretation. Therefore, before the deepest encoder feature is propagated through the pyramid decoding pathway, its contextual representation should be enriched in an input-adaptive manner.

Conventional pyramid pooling aggregates contextual information at several predefined scales using fusion parameters shared across all input samples [52]. Although the branch features remain input-dependent, the module does not explicitly estimate the preference of each input image for different contextual scales. This limitation is relevant to SAR oil spill segmentation because different scenes may require different contextual ranges: large-area slicks may benefit from a coarse global context, whereas fragmented slicks and small ambiguous dark regions may require a finer regional context. To address this issue, we introduce the Dynamic Scale Pyramid Context (DSPC) module at the deepest encoder stage C_4 , as illustrated in Figure 3.

Let $x = C_4 \in \mathbb{R}^{B \times D_4 \times H_4 \times W_4}$ denote the deepest encoder feature, where $D_4 = 768$ for ConvNeXtV2-Tiny. DSPC contains $N = 4$ contextual branches associated with the pooling scales $(s_1, s_2, s_3, s_4) = (1, 2, 3, 6)$. The i -th branch first performs adaptive average pooling to a spatial size of $s_i \times s_i$, projects the pooled feature to the decoder dimension $C_d = 256$, and then bilinearly resizes it to the spatial resolution of x :

$$B_i(x) = \text{Up}_{H_4, W_4} \left(\rho \left(\text{Conv}_{1 \times 1}^{D_4 \rightarrow C_d} (\text{AdaPool}_{s_i}(x)) \right) \right), \quad i = 1, \dots, N, \quad (5)$$

where $\rho(\cdot) = \text{ReLU}(\text{BN}(\cdot))$, $\text{Up}_{H_4, W_4}(\cdot)$ denotes bilinear resizing to $H_4 \times W_4$, and $B_i(x) \in \mathbb{R}^{B \times C_d \times H_4 \times W_4}$.

To make the relative contribution of each contextual scale input-adaptive, DSPC generates branch-level logits from the global representation of x . Global average pooling first compresses the spatial dimensions of x , after which a lightweight two-layer gating function reduces the channel dimension from D_4 to $D_4/16 = 48$ and maps the resulting representation to N scale logits:

$$z = \text{Conv}_{1 \times 1}^{D_4/16 \rightarrow N} \left(\text{ReLU} \left(\text{Conv}_{1 \times 1}^{D_4 \rightarrow D_4/16} (\text{GAP}(x)) \right) \right), \quad z \in \mathbb{R}^{B \times N \times 1 \times 1}. \quad (6)$$

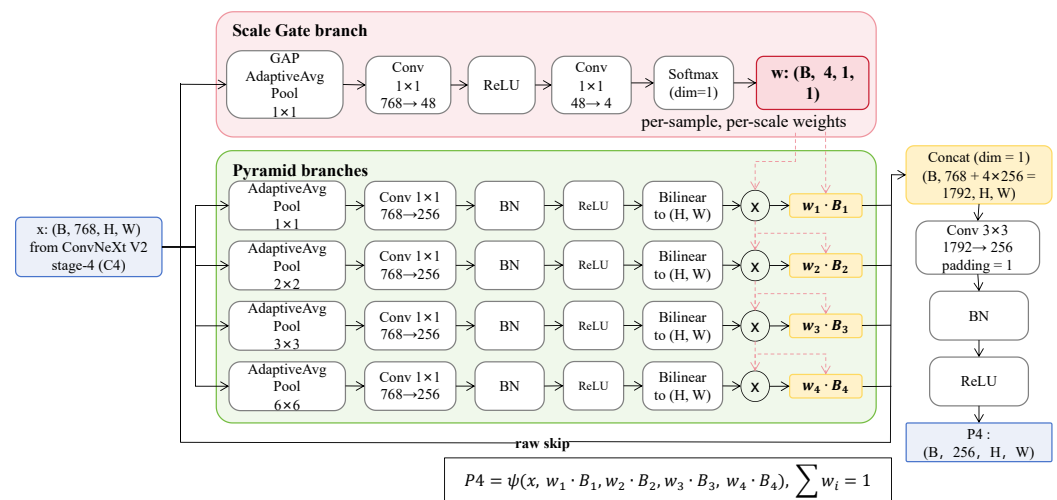


Figure 3. Structure of the proposed Dynamic Scale Pyramid Context (DSPC) module. Given the deepest encoder feature $x = C_4$, DSPC constructs four contextual branches with pooling scales $(s_1, s_2, s_3, s_4) = (1, 2, 3, 6)$. An input-conditioned softmax scale gate assigns a per-sample weight to each branch. The weighted contextual features are concatenated with the unpooled feature x and projected through a 3×3 bottleneck convolution to generate P_4 . Here, H_4 and W_4 denote the spatial dimensions of the C_4 feature.

The branch weights are obtained by applying softmax along the scale dimension:

$$w_i = \frac{\exp(z_i)}{\sum_{j=1}^N \exp(z_j)}, \quad i = 1, \dots, N, \quad (7)$$

where $w \in \mathbb{R}^{B \times N \times 1 \times 1}$ and $\sum_{i=1}^N w_i = 1$. Each w_i is a per-sample scalar weight that is broadcast over all channels and spatial locations of the corresponding contextual branch.

The dynamically weighted contextual branches are concatenated with the original unpooled feature x along the channel dimension and compressed by a 3×3 bottleneck convolution:

$$P_4 = \text{DSPC}(x) = \rho \left(\text{Conv}_{3 \times 3}^{D_4 + NC_d \rightarrow C_d} (\text{Concat}_c [x, w_1 B_1(x), \dots, w_N B_N(x)]) \right), \quad (8)$$

where $\text{Concat}_c(\cdot)$ denotes channel-wise concatenation. Since $D_4 + NC_d = 768 + 4 \times 256 = 1792$, the bottleneck convolution maps the concatenated feature from 1792 channels to $C_d = 256$ channels using a stride of one and padding of one. The resulting feature satisfies $P_4 \in \mathbb{R}^{B \times C_d \times H_4 \times W_4}$.

The application of unpooled feature x is justified by the fact that multiple applications of spatial pooling may cause dense features to become ineffective for boundary localization and class discrimination. This is why bottleneck projection integrates both original high-level representations and contextual information extracted at different spatial levels.

DSPC uses softmax to assign four non-negative scale weights that sum to one for each input image. Increasing the relative weight of one scale therefore reduces the combined weight of the others. Independent sigmoid gates assign separate branch weights without imposing this constraint. DSPC also differs from SKNet-style selection, which performs softmax-based selection among convolutional kernels and commonly generates channel-dependent weights [53]. Here, the competition is performed among pooled contextual branches at the deepest semantic stage using one branch-level weight vector for each input image. Standard ASPP-type modules, by comparison, generally concatenate or fuse parallel dilation and pooling branches without explicitly normalizing their relative contributions over the scale dimension.

DSPC employs a per-image scale gate rather than a spatially varying gate. Since the module operates on the lowest-resolution encoder feature, its purpose is to estimate the dominant contextual requirement of the input scene before top-down decoding, rather than to perform fine-grained regional modulation. A spatially varying scale gate would introduce additional memory and computational costs, while spatial adaptation is subsequently handled by the content-adaptive and boundary-sensitive fusion operations in EDG-FPN.

The computational overhead of DSPC is controlled by applying it only to the deepest and lowest-resolution feature. The scale gating branch reduces the channel dimension from D_4 to $D_4/16$ before predicting the N branch weights, and each pooled feature is projected to the decoder dimension before fusion. The unpooled branch preserves dense high-level semantics, while the softmax gate enforces normalized competition among contextual scales.

2.4. Edge-Aware Dynamic Gated FPN Decoder

After SDLE refines the shallow-to-mid-level lateral features and DSPC adapts the deepest contextual representation, the remaining problem is how to fuse multilevel features into a prediction that is both spatially precise and semantically reliable. Conventional pyramid decoders commonly rely on unconditional top-down addition, in which deep features are upsampled using fixed interpolation [54]. Although computationally efficient, this content-agnostic strategy may propagate unreliable semantic responses and oversmooth weak oil film boundaries. This issue is particularly important for SAR oil spill segmentation because the transitions among oil films, the sea surface, and look-alike dark formations are often gradual and unstable.

To address this issue, we propose the Edge-Aware Dynamic Gated Feature Pyramid Network (EDG-FPN) decoder, as illustrated in Figure 4. EDG-FPN contains two types of fusion units. First, DySample-based dynamic gated fusion is employed at the deeper top-down stages to adaptively control the propagation of high-level semantic information. Second, a boundary-aware gate (BAG) is introduced at the highest-resolution fusion stage, where a lightweight EdgeHead predicts a boundary prior from the finest lateral feature and uses it to modulate the final top-down fusion. Let the four channel-aligned decoder inputs be denoted by $\{L_1, L_2, L_3, P_4\}$, where L_1 is the lateral feature derived from C_1 ; L_2 and L_3 are the SDLE-refined lateral features derived from C_2 and C_3 , respectively; and P_4 is the DSPC-enhanced deepest feature. All features have the unified decoder width $C_d = 256$.

For the two deeper fusion stages, namely $P_4 \rightarrow L_3$ and $P_3 \rightarrow L_2$, EDG-FPN employs a DySample-based dynamic gated fusion unit. At each level $i \in \{3, 2\}$, the deeper pyramid feature P_{i+1} is first upsampled to the spatial resolution of the corresponding lateral feature L_i using DySample [48]:

$$\hat{P}_{i+1} = \text{DySample}(P_{i+1}). \quad (9)$$

A content-based fusion gate is then generated from the channel-wise concatenation of L_i and $\hat{P}_{i+1}$:

$$G_i = \sigma \left(\text{Conv}_{1 \times 1}^{2C_d \rightarrow C_d} \left(\text{Concat}_c \left[L_i, \hat{P}_{i+1} \right] \right) \right), \quad (10)$$

where $\text{Concat}_c(\cdot)$ denotes channel-wise concatenation, $\sigma(\cdot)$ denotes the sigmoid function, and $G_i \in [0, 1]^{B \times C_d \times H_i \times W_i}$ is a spatially and channel-wise varying fusion gate.

The fused feature at level i is computed as

$$P_i = L_i + G_i \odot \hat{P}_{i+1}, \quad (11)$$

where $\odot$ denotes element-wise multiplication. This formulation allows the decoder to determine where and through which channels high-level semantic information should be introduced, rather than uniformly adding the upsampled deep feature to all lateral responses.

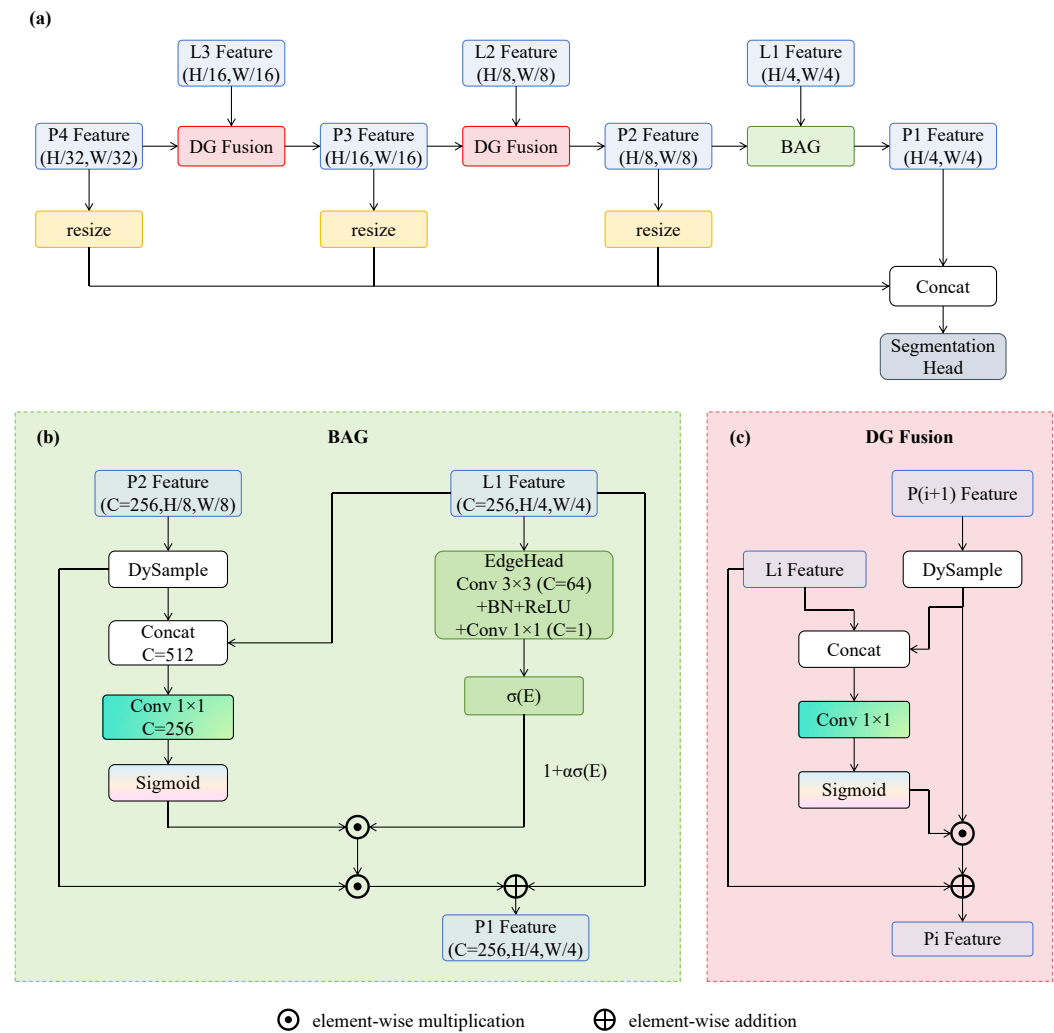


Figure 4. Structure of the proposed EDG-FPN decoder. (a) Overall top-down fusion and multilevel feature aggregation. (b) Boundary-aware gate (BAG), where EdgeHead predicts the boundary prior $\sigma(E)$ from L_1 and uses $1 + \alpha\sigma(E)$ to modulate the content-based fusion coefficient. (c) DySample-based dynamic gated fusion used at the deeper pyramid levels.

At the highest-resolution fusion stage $P_2 \rightarrow L_1$, EDG-FPN replaces the ordinary dynamic gated fusion with the proposed BAG. Since L_1 has the highest spatial resolution among the lateral features, it retains relatively detailed local boundary information. A lightweight EdgeHead is attached to L_1 to predict a single-channel boundary logit map:

$$E = \text{Conv}_{1 \times 1}^{64 \rightarrow 1} \left(\text{ReLU} \left(\text{BN} \left(\text{Conv}_{3 \times 3}^{C_d \rightarrow 64} (L_1) \right) \right) \right), \quad E \in \mathbb{R}^{B \times 1 \times H_1 \times W_1}. \quad (12)$$

The hidden channel width of EdgeHead is set to 64. The boundary logits E are converted into a boundary probability map $\sigma(E)$, which is used as an internal fusion prior rather than as an independent final prediction. EdgeHead is supervised using boundary targets derived directly from the segmentation labels during training and therefore requires no additional manual edge annotations.

Let

$$\hat{P}_2 = \text{DySample}(P_2) \quad (13)$$

denote the upsampled P_2 feature. The BAG first computes a content-based gate using the same formulation as the deeper fusion units:

$$G_1^{\text{base}} = \sigma \left(\text{Conv}_{1 \times 1}^{2C_d \rightarrow C_d} \left(\text{Concat}_c \left[L_1, \hat{P}_2 \right] \right) \right). \quad (14)$$

The base gate is then modulated by the predicted boundary prior:

$$G_1^{\text{bag}} = G_1^{\text{base}} \odot (\mathbf{1} + \alpha \sigma(E)), \quad (15)$$

where α is a learnable scalar initialized to 0.5, $\mathbf{1}$ denotes an all-one tensor with a broadcast-compatible shape, and $\sigma(E) \in [0, 1]^{B \times 1 \times H_1 \times W_1}$ is broadcast along the channel dimension. The resulting G_1^{bag} is a boundary-modulated fusion coefficient rather than a strictly sigmoid-bounded gate.

The highest-resolution pyramid feature is subsequently obtained as

$$P_1 = L_1 + G_1^{\text{bag}} \odot \hat{P}_2. \quad (16)$$

For pixels far from semantic boundaries, $\sigma(E)$ tends to be small and the boundary-modulated coefficient approaches the content-based gate G_1^{base} . In boundary-sensitive regions, the predicted boundary prior modulates the contribution of the upsampled high-level feature. Since this modulation is applied multiplicatively to the content-based coefficient, the boundary prior adjusts an existing content-consistent fusion response rather than replacing it.

After top-down fusion, P_2 , P_3 , and P_4 are bilinearly resized to the spatial resolution of P_1 . The four pyramid features are then concatenated along the channel dimension and compressed by a 3×3 bottleneck convolution:

$$F_{\text{agg}} = \text{ReLU} \left(\text{BN} \left(\text{Conv}_{3 \times 3}^{4C_d \rightarrow C_d} (\text{Concat}_c [P_1, P'_2, P'_3, P'_4]) \right) \right), \quad (17)$$

where P'_i denotes P_i resized to the spatial resolution of P_1 . Since each pyramid feature contains $C_d = 256$ channels, the concatenated feature contains $4C_d = 1024$ channels and is compressed to $C_d = 256$ channels.

A dropout layer and a 1×1 classifier are subsequently applied to produce the segmentation logits, which are bilinearly upsampled to the input resolution:

$$Y = \text{Up}_{H,W} \left(\text{Conv}_{1 \times 1}^{C_d \rightarrow K} (\text{Dropout}(F_{\text{agg}})) \right), \quad Y \in \mathbb{R}^{B \times K \times H \times W}, \quad (18)$$

where $\text{Up}_{H,W}(\cdot)$ denotes bilinear resizing to the input spatial resolution and $K = 5$ is the number of semantic categories.

2.5. Training Objective

The network is trained using pixel-wise segmentation supervision together with label-derived boundary supervision for EdgeHead. Let $\hat{Y}$ denote the ground-truth segmentation mask. The segmentation objective combines the class-weighted cross-entropy and per-class soft Dice loss:

$$\mathcal{L}_{\text{seg}}(Y, \hat{Y}) = \mathcal{L}_{\text{CE}}(Y, \hat{Y}; \mathbf{w}) + \lambda_{\text{dice}} \mathcal{L}_{\text{Dice}}(Y, \hat{Y}), \quad (19)$$

where $\mathbf{w}$ contains the class weights used to account for the unequal numbers of pixels among the OSDD classes.

To supervise EdgeHead without requiring additional boundary annotations, a boundary supervision map T is generated online from $\hat{Y}$ during training. Specifically, the segmentation mask is one-hot-encoded and processed using class-wise depth-wise Sobel filtering. The resulting gradient responses are aggregated across classes by a maximum operation, binarized, and slightly dilated to alleviate boundary sparsity. The boundary map is then bilinearly resized to the spatial resolution of the predicted edge logits E , and the interpolated values are treated as soft boundary targets. A validity mask V is generated in parallel

to exclude pixels carrying the ignore label. This target generation procedure contains no trainable parameters and is performed without gradient computation.

EdgeHead is supervised using a combination of binary cross-entropy and the soft Dice loss:

$$\mathcal{L}_{\text{edge}} = \mathcal{L}_{\text{BCE}}(\sigma(E), T; V) + \mathcal{L}_{\text{Dice}}(\sigma(E), T; V), \quad (20)$$

where $\sigma(\cdot)$ denotes the sigmoid function.

The complete training objective is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}}(Y, \hat{Y}) + \lambda_{\text{edge}} \mathcal{L}_{\text{edge}}, \quad (21)$$

where λ_{edge} controls the contribution of the boundary supervision term. The values of $\mathbf{w}$, λ_{dice} , and λ_{edge} are reported in Section 3.2. Through the coupling between E and the boundary-aware gate, EdgeHead receives direct supervision from $\mathcal{L}_{\text{edge}}$ as well as indirect segmentation gradients from $\mathcal{L}_{\text{seg}}$, encouraging the predicted boundary prior to represent semantic transitions that are useful for the final segmentation task.

Overall, EDG-FPN improves multilevel feature fusion through complementary content-adaptive and boundary-sensitive mechanisms. At the deeper pyramid stages, the DySample-based fusion units adjust the contribution of upsampled features using weights predicted from the input features. At the highest-resolution stage, the BAG uses the boundary map predicted by EdgeHead, which is trained using targets generated from segmentation labels. Together with SDLE and DSPC, EDG-FPN enables SDE-Net to jointly model elongated local structures, an input-adaptive high-level context, and boundary-sensitive cross-level feature fusion within a unified segmentation framework.

3. Experiments and Results

3.1. Dataset and Evaluation Metrics

The experiments are conducted on the public Oil Spill Detection Dataset (OSDD) introduced by Krestenitis et al. [15]. The OSDD was developed for pixel-level SAR oil spill interpretation using Sentinel-1 imagery acquired through the Copernicus Open Access Hub. The geographic coordinates and timestamps of confirmed pollution events were provided by the European Maritime Safety Agency (EMSA) through the CleanSeaNet service. The dataset providers processed Sentinel-1 VV-polarization data by cropping regions of interest and resizing them to 1250×650 pixels, applying radiometric calibration and a 7×7 median filter for speckle suppression, and converting the data to image brightness values [15]. These operations were performed during dataset construction; the model input processing used in this study is described in Section 3.2.

Each pixel in the OSDD is annotated to one of five mutually exclusive semantic categories: sea surface, oil spill, look-alike, ship, and land. The sea surface class serves as the background, while oil spill is the primary target class. The look-alike class denotes non-oil dark formations that share similar low-backscatter signatures to real oil slicks and therefore constitutes the main source of false alarms. The ship and land classes are retained because they provide additional contextual information for multiclass maritime interpretation. The ship class represents extremely sparse small targets in maritime scenes, while coastal land areas and near-shore structures may increase the ambiguity of dark region interpretation.

Representative SAR images and their corresponding pixel-level annotations are shown in Figure 5. These examples highlight the visual ambiguity between oil spills and look-alike dark formations, as well as the dominance of the sea surface background in SAR oil spill segmentation scenes.

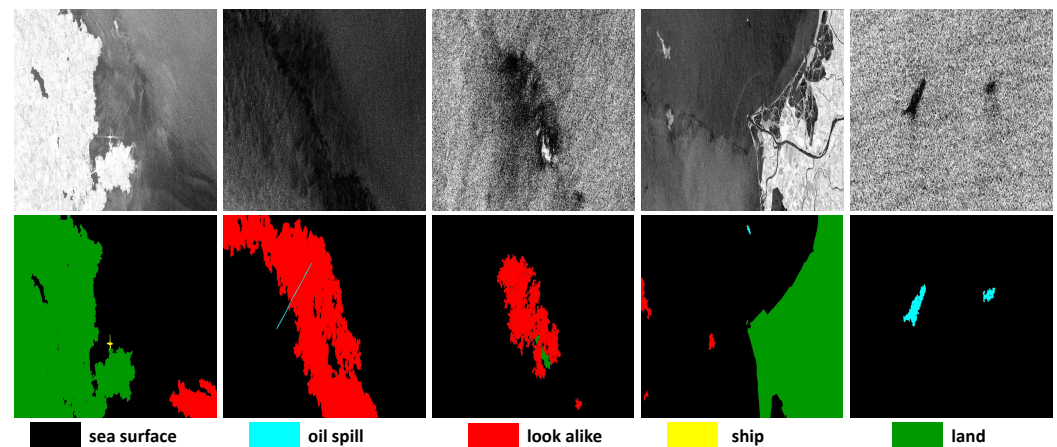


Figure 5. Representative samples from the Oil Spill Detection Dataset. The first row shows SAR intensity images, and the second row shows the corresponding pixel-level annotations. Black, cyan, red, yellow, and green denote sea surface, oil spill, look-alike, ship, and land, respectively. The colors are used only for visualization; training uses integer class labels (0–4).

Following the official split of Krestenitis et al. [15], the OSDD contains 1002 training images and 110 test images, of which 88 contain oil spill pixels. Our count of the original-resolution test masks yields 234 annotated oil spill regions, defined as 8-connected components of label 1 without size filtering. These image-level counts do not establish the numbers of independent SAR scenes or spill events, because the released data lack source-scene and event identifiers. In our experiments, the official training split is further divided into a training subset and a validation subset for model optimization and checkpoint selection. As a result, the training, validation, and testing subsets approximately follow an 8:1:1 ratio. The official test split is used only for final evaluation and is not involved in model selection or hyperparameter tuning.

Table 1 summarizes the label identifier, display color, and per-class pixel statistics computed from the official training and test masks before the internal training–validation partition. The dataset exhibits severe pixel-level imbalance: the sea surface class accounts for about 88% of the pixels in the official training split, whereas oil spill and look-alike together occupy less than 7%, and ship pixels are extremely sparse. This distribution makes the OSDD challenging for semantic segmentation because a model may easily favor the dominant background class while missing rare but operationally important categories.

Table 1. Class definitions and per-class pixel statistics of the official OSDD training and test splits before the internal training–validation partition. Pixel counts are reported in millions (M), and the proportions may not sum to 100% because of rounding.

Category	Color	Label	Pixels (M)		Proportion (%)	
			Train	Test	Train	Test
Sea	Black	0	716.89	80.90	88.06	90.52
Oil Spill	Cyan	1	8.19	0.96	1.01	1.07
Look-Alike	Red	2	45.83	4.63	5.63	5.18
Ship	Yellow	3	0.35	0.01	0.04	0.01
Land	Green	4	42.87	2.88	5.27	3.22

Overall, the OSDD contains the main difficulties targeted in this study. First, oil slicks vary in shape and spatial extent, appearing as elongated streaks, compact patches, or fragmented regions. Some slicks have low contrast against the surrounding sea surface and poorly defined boundaries. Second, look-alike phenomena can produce similar dark

formations, making both the target morphology and surrounding context relevant to interpretation [15]. Third, ship instances are extremely sparse and occupy only small image areas, making their representations difficult to learn under the severely imbalanced pixel distribution. These characteristics make the OSDD a suitable benchmark for evaluating shallow strip-structured refinement, adaptive high-level context aggregation, and boundary-aware dynamic gated feature fusion.

For the quantitative evaluation, we report the mean intersection over union (mIoU) and mean Dice coefficient (mDice), which are commonly used in semantic segmentation. For each class k , let TP_k , FP_k , and FN_k denote the numbers of true-positive, false-positive, and false-negative pixels, respectively. The per-class IoU and Dice coefficient are defined as

$$\text{IoU}_k = \frac{TP_k}{TP_k + FP_k + FN_k}, \quad \text{Dice}_k = \frac{2TP_k}{2TP_k + FP_k + FN_k}. \quad (22)$$

The final mIoU and mDice are obtained by averaging over all $K = 5$ semantic categories:

$$\text{mIoU} = \frac{1}{K} \sum_{k=1}^K \text{IoU}_k, \quad \text{mDice} = \frac{1}{K} \sum_{k=1}^K \text{Dice}_k. \quad (23)$$

Both the mIoU and mDice are averaged over all five semantic categories, including the sea surface background class. Pixels assigned the ignore label are excluded from the computation of all segmentation metrics. In addition to the mIoU and mDice, the per-class IoU is reported for sea surface, oil spill, look-alike, ship, and land to examine the class-wise behavior of each method. In this paper, the mIoU is used as the primary evaluation metric, while the mDice is reported as a complementary region overlap measure. Compared with the overall pixel accuracy, these two metrics are more suitable for the OSDD because they reduce the dominance of the large sea surface background and better reflect the segmentation quality of rare classes such as oil spill, look-alike, and ship.

3.2. Implementation Details

All experiments are implemented in PyTorch (version 2.6.0) and conducted on a workstation equipped with two NVIDIA GeForce RTX 4090 GPUs (NVIDIA Corporation, Santa Clara, CA, USA). The ConvNeXtV2-Tiny backbone is initialized from a publicly released ImageNet-22K-pretrained checkpoint. Batch normalization is used in the decoder components, and all lateral features are projected to a unified decoder width of $C_d = 256$. A dropout layer with a ratio of 0.15 is applied before the main segmentation classifier. No auxiliary FCN segmentation head is used in the final model. EdgeHead remains active during both training and inference because its predicted boundary prior is used to modulate the boundary-aware gate.

Each single-channel SAR image is replicated into three channels to match the input format of the ImageNet-pretrained backbone and is normalized using the ImageNet channel mean and standard deviation. During training, each image and its corresponding mask are jointly augmented using random scaling, random cropping to 512×512 , random horizontal and vertical flipping, and random rotations selected from $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$. All compared methods are used with the same input preprocessing, data augmentation, dataset partition, and evaluation protocol to ensure a fair comparison.

Loss functions can help to address class imbalance, as illustrated by the focal binary cross-entropy used by Parasiris et al. for rare event prediction [55]. In this study, we address the pixel-level imbalance in SAR oil spill segmentation by computing frequency-based cross-entropy class weights exclusively from valid pixels in the OSDD training subset. Let n_c denote the number of training pixels belonging to class c , and let p_c denote the corresponding pixel frequency:

$$p_c = \frac{n_c}{\sum_{j=1}^K n_j}. \quad (24)$$

Direct inverse-frequency weighting is avoided because the extremely sparse ship class would otherwise receive an excessively large weight and could dominate optimization. We therefore employ a logarithmically compressed inverse-frequency score:

$$\tilde{w}_c = \frac{1}{\ln(\kappa + p_c)}, \quad (25)$$

where $\kappa = 1.05$ is a smoothing constant that limits the weight variation caused by extremely small class frequencies. The scores are subsequently bounded and rescaled as

$$w_c = \max\left(w_{\min}, w_{\max} \frac{\tilde{w}_c}{\max_j \tilde{w}_j}\right), \quad (26)$$

where $w_{\min} = 0.5$ and $w_{\max} = 3.0$. This transformation compresses the extreme class frequency differences, prevents the rare ship class from producing excessively large gradients, and retains sufficient supervision for the frequent background categories. After rounding to two decimal places, the resulting class weight vector is $\mathbf{w} = [0.50, 2.53, 1.46, 3.00, 1.51]$ for sea surface, oil spill, look-alike, ship, and land, respectively.

The Dice loss coefficient is empirically set to $\lambda_{\text{dice}} = 2.0$ and fixed for all experiments. This coefficient provides additional region overlap supervision alongside the class-weighted cross-entropy loss. The boundary loss coefficient is selected through a coarse search on the validation subset while keeping $\mathbf{w}$ and λ_{dice} fixed. Among the evaluated candidates, $\lambda_{\text{edge}} = 0.4$ achieves the highest validation mIoU and is therefore used in all subsequent experiments. The online boundary target generation procedure follows Section 2.4, and the same boundary supervision configuration is retained throughout all experiments. Pixels carrying the ignore label are excluded from both the segmentation and boundary losses.

The model is optimized using AdamW [56] with a base learning rate of 6×10^{-5} and weight decay of 0.01. Layer-wise learning rate decay with a factor of 0.91 is applied to the ConvNeXtV2-Tiny backbone. The learning rate is linearly warmed up during the first 15 epochs and subsequently decayed using a polynomial schedule with power of 0.9. The network is trained for 80 epochs with a batch size of 8 per GPU, resulting in a global batch size of 16. Gradient clipping with an ℓ_2 -norm threshold of 1.0 is applied, and automatic mixed-precision training is enabled to reduce memory consumption.

The checkpoint achieving the highest validation mIoU is selected for final evaluation on the official test split. During inference, the predicted logits are bilinearly resized to the original image resolution before metric computation. The parameter count and FLOPs are calculated from the inference graph using an input tensor of size $3 \times 512 \times 512$. EdgeHead is included in the complexity calculation because it remains active during inference.

3.3. Comparison with Representative Methods

To evaluate the effectiveness of SDE-Net, we compare it with representative semantic segmentation architectures, including U-Net [29], DeepLabV3+ [57], SegFormer [58], UPer-Net [47], SegNeXt [59], and FCS-Net [39]. These methods cover conventional convolutional encoder–decoder networks, atrous convolution-based architectures, Transformer-based models, and hierarchical pyramid decoders. All methods are evaluated using the same dataset partition, input preprocessing, data augmentation strategy, and metric computation protocol. The backbone configurations used for comparison are listed in Table 2,

and the FCS-Net backbone adaptation is described in Section 4. The quantitative results are reported in Table 2.

Table 2. Quantitative comparison with representative semantic segmentation methods on OSDD. All segmentation scores are reported as percentages. Bold and underlined values denote the best and second-best results, respectively. Higher values are better for segmentation metrics, whereas lower values are better for Params and FLOPs.

Method	Backbone	mIoU ↑	mDice ↑	Sea IoU ↑	Oil IoU ↑	Look-Alike IoU ↑	Ship IoU ↑	Land IoU ↑	Params (M) ↓	FLOPs (G) ↓
U-Net	ResNet-50	61.58	72.96	94.86	48.76	39.34	32.95	92.00	47.13	197.85
DeepLabV3+	ResNet-50	65.54	76.16	96.57	49.07	60.04	29.49	92.55	41.22	176.50
SegFormer	MiT-B3	64.19	74.35	<u>96.74</u>	47.58	60.54	22.63	93.46	44.60	41.93
UPerNet	Swin-T	66.93	77.65	96.73	48.75	61.53	35.33	92.29	58.94	235.97
SegNeXt	MSCAN-L	67.21	77.75	96.66	49.59	60.19	34.84	94.76	48.78	<u>69.23</u>
UPerNet	ConvNeXtV2-T	68.69	79.04	96.82	52.73	<u>61.37</u>	36.93	<u>95.61</u>	<u>37.09</u>	76.57
FCS-Net	ConvNeXtV2-T	<u>70.04</u>	<u>80.68</u>	96.24	<u>55.32</u>	54.15	49.90	94.61	40.17	97.86
SDE-Net	ConvNeXtV2-T	72.94	82.88	96.41	61.64	60.54	<u>49.61</u>	96.50	36.52	69.55

Note: ↑ indicates that higher values are better; ↓ indicates that lower values are better.

As shown in Table 2, SDE-Net achieves the best overall segmentation performance, with an mIoU of 72.94% and mDice of 82.88%. Compared with the second-best FCS-Net, SDE-Net improves the mIoU and mDice by 2.90 and 2.20 percentage points, respectively. Moreover, compared with the UPerNet baseline equipped with the same ConvNeXtV2-Tiny backbone, SDE-Net improves the mIoU from 68.69% to 72.94% and the mDice from 79.04% to 82.88%, corresponding to gains of 4.25 and 3.84 percentage points. Since both models employ the same encoder and are evaluated under the same experimental protocol, the performance improvement mainly demonstrates the effectiveness of the proposed decoder redesign and task-oriented feature enhancement modules.

The improvement is particularly significant for the primary oil spill category. SDE-Net achieves an Oil IoU of 61.64%, outperforming the same-backbone UPerNet baseline and FCS-Net by 8.91 and 6.32 percentage points, respectively. This result indicates that the proposed directional refinement, adaptive contextual aggregation, and dynamic feature fusion strategies improve the representation and delineation of irregular oil spill regions. In addition, SDE-Net achieves a Ship IoU of 49.61%, increasing the value of the same-backbone UPerNet baseline from 36.93% by 12.68 percentage points. Although FCS-Net obtains a slightly higher Ship IoU of 49.90%, SDE-Net provides substantially better overall performance and a significantly higher Oil IoU.

For the dominant sea surface category, SDE-Net achieves an IoU of 96.41%, which remains competitive with the best result of 96.82%. Its Look-Alike IoU reaches 60.54%, slightly lower than that of the same-backbone UPerNet baseline (61.37%). This indicates that the overall improvement of SDE-Net does not result from uniformly increasing all category scores but mainly from improving the segmentation of challenging foreground categories, especially oil spill and ship. The remaining confusion between oil spill and look-alike regions also reflects the inherent difficulty caused by their similar low-backscatter characteristics in SAR imagery.

In terms of computational efficiency, SDE-Net achieves a favorable accuracy–complexity trade-off. The proposed model contains 36.52M parameters, which is lower than in the other compared methods, and requires 69.55G FLOPs. Compared with the UPerNet-ConvNeXtV2-T baseline, SDE-Net reduces the parameter count by 0.57M and decreases the computational cost by 7.02G FLOPs while achieving substantially higher segmentation accuracy. Although SegFormer obtains the lowest FLOPs, its mIoU is 8.75 percentage points lower than that of SDE-Net. SegNeXt has a similar computational

cost to SDE-Net but achieves a lower mIoU by 5.73 percentage points. These results demonstrate that SDE-Net provides an effective balance between segmentation accuracy and computational efficiency.

Figure 6 presents a qualitative comparison between SDE-Net and some representative segmentation methods on challenging SAR oil spill scenes. The selected examples cover elongated oil slicks, fragmented structures, oil spill/look-alike confusion, sparse targets, and complex coastal backgrounds.

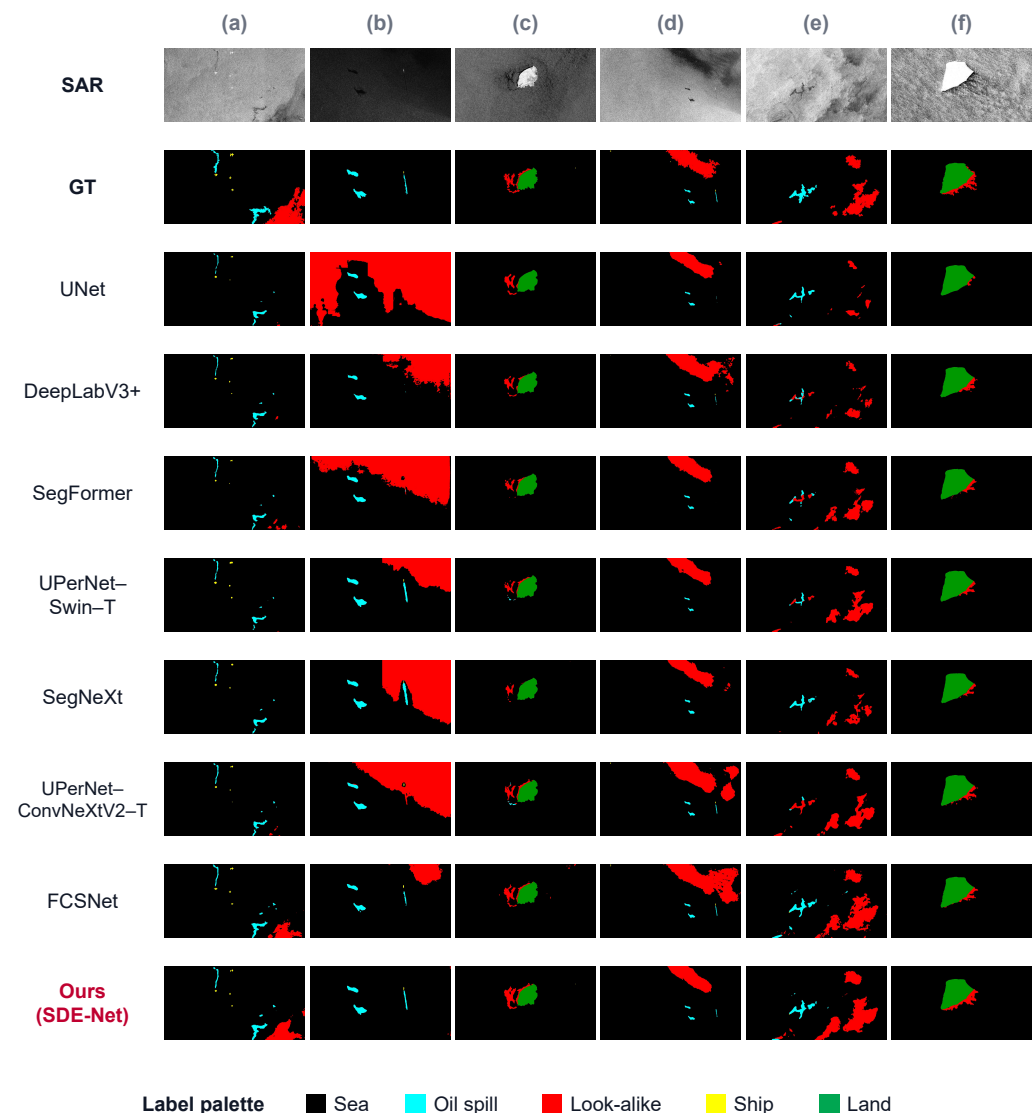


Figure 6. Qualitative comparison on representative OSDD test images. Columns (a–f) show six test examples. From top to bottom, the rows present the SAR image, ground truth (GT), U-Net, DeepLabV3+, SegFormer, UPerNet with Swin-T, SegNeXt, UPerNet with ConvNeXtV2-T, FCS-Net, and the proposed SDE-Net. Black, cyan, red, yellow, and green denote sea surface, oil spill, look-alike, ship, and land, respectively.

In scenes containing elongated and fragmented oil films, SDE-Net produces more continuous and complete predictions than the comparison methods. Conventional segmentation networks tend to lose narrow structures or generate discontinuous responses due to the weak contrast and anisotropic morphology of oil slicks, whereas SDE-Net preserves more detailed structural information. This observation demonstrates the effectiveness of directional local feature enhancement in capturing elongated oil spill patterns.

For scenes containing closely distributed oil spill and look-alike regions, several comparison methods produce inaccurate responses in ambiguous dark areas or fail to recover complete oil spill regions. In contrast, SDE-Net provides more consistent separation between these visually similar categories by combining adaptive contextual representation with boundary-aware feature fusion.

Moreover, SDE-Net shows the improved preservation of sparse small targets and more stable predictions near coastal regions. Compared with other methods, it reduces fragmented responses and maintains more complete target contours. These qualitative observations are consistent with the quantitative improvements in the Oil IoU, Ship IoU, and Land IoU reported in Table 2.

Overall, the qualitative comparison further verifies that SDE-Net can effectively address the structural, contextual, and boundary-related challenges in SAR oil spill segmentation. The proposed SDLE, DSPC, and EDG-FPN modules jointly improve the representation of complex oil spill regions while maintaining competitive computational efficiency.

3.4. Ablation Studies

To evaluate the individual contributions and complementarity of the proposed components, we conduct ablation experiments using UPerNet with the ConvNeXtV2-Tiny backbone as the baseline. SDLE, DSPC, and EDG-FPN are introduced individually and in different combinations, while all variants follow the same dataset partition, training configuration, and evaluation protocol described in Section 3.2. After fixing the final architecture according to the validation results, the ablation variants are retrained and evaluated on the official test split. The results are reported in Table 3.

Table 3. Ablation study of the main components on the OSDD test split. All segmentation scores are reported as percentages. Bold and underlined values denote the best and second-best results, respectively. Higher values are better for segmentation metrics, whereas lower values are better for Params and FLOPs. All variants use fixed class weights $\mathbf{w} = [0.50, 2.53, 1.46, 3.00, 1.51]$ for sea surface, oil spill, look-alike, ship, and land, respectively, and a fixed Dice loss coefficient $\lambda_{\text{dice}} = 2.0$.

SDLE	DSPC	EDG-FPN	mIoU $\uparrow$	mDice $\uparrow$	Sea $\uparrow$	Oil $\uparrow$	Look-Alike $\uparrow$	Ship $\uparrow$	Land $\uparrow$	Params (M) $\downarrow$	FLOPs (G) $\downarrow$
–	–	–	68.69	79.04	<u>96.82</u>	52.73	61.37	36.93	95.61	37.09	76.57
✓	–	–	70.46	80.94	96.44	55.94	58.89	46.44	94.56	37.69	76.96
–	✓	–	70.88	81.32	96.85	55.05	<u>60.57</u>	47.99	93.95	37.57	76.57
–	–	✓	69.75	80.29	96.26	55.98	56.47	44.74	95.32	36.33	69.16
✓	✓	–	<u>71.14</u>	<u>81.44</u>	96.64	<u>56.15</u>	58.79	48.13	96.01	37.72	76.96
✓	–	✓	70.90	81.23	96.40	55.56	58.08	48.02	<u>96.45</u>	36.48	<u>69.55</u>
–	✓	✓	71.05	81.38	96.36	55.78	58.32	<u>48.64</u>	96.13	<u>36.37</u>	69.16
✓	✓	✓	72.94	82.88	96.41	61.64	60.54	49.61	96.50	36.52	<u>69.55</u>

Note: $\uparrow$ indicates that higher values are better; $\downarrow$ indicates that lower values are better.

As shown in Table 3, each proposed component improves the overall segmentation performance of the baseline. Introducing SDLE alone increases the mIoU from 68.69% to 70.46% and the mDice from 79.04% to 80.94%. The Oil IoU and Ship IoU increase by 3.21 and 9.51 percentage points, respectively, indicating that directional local refinement benefits elongated oil film structures and sparse small targets. The reduction in the Look-Alike IoU, however, suggests that shallow directional enhancement alone cannot fully resolve the ambiguity between oil spills and visually similar dark formations.

Introducing DSPC alone produces the highest overall performance among the single-module variants, reaching a 70.88% mIoU and 81.32% mDice. It also increases the Ship IoU from 36.93% to 47.99% while maintaining a Look-Alike IoU of 60.57%, which is the best

value among the single-module configurations. These results support the use of DSPC to combine context at different scales using weights predicted for each input image.

EDG-FPN alone increases the mIoU to 69.75%, with the Oil IoU and Ship IoU improving to 55.98% and 44.74%, respectively. Although its accuracy gain is smaller than those of SDLE and DSPC when used independently, EDG-FPN reduces the parameter count from 37.09M to 36.33M and the FLOPs from 76.57G to 69.16G. This result demonstrates that the redesigned cross-level fusion pathway improves the segmentation performance while reducing the computational cost of the original UPerNet decoder.

The pairwise configurations further demonstrate the complementarity of the proposed modules. The combinations SDLE+DSPC, SDLE+EDG-FPN, and DSPC+EDG-FPN achieve mIoU values of 71.14%, 70.90%, and 71.05%, respectively, all exceeding those of the corresponding single-module configurations. Among them, SDLE+DSPC provides the strongest pairwise result, whereas configurations containing EDG-FPN retain lower model complexity and computational costs.

The complete SDE-Net achieves the best overall performance, with an mIoU of 72.94% and an mDice of 82.88%. Compared with the baseline, the full model improves the mIoU and mDice by 4.25 and 3.84 percentage points, respectively. The improvements are particularly pronounced for the primary oil spill category and the extremely sparse ship category, whose IoU values increase by 8.91 and 12.68 percentage points. The Land IoU also increases from 95.61% to 96.50%, while the Sea IoU remains stable at 96.41%. Although the Look-Alike IoU is 0.83 percentage points lower than that of the baseline, the complete model recovers most of the degradation observed in several incomplete configurations and achieves substantially better class-averaged performance. In addition, SDE-Net uses 36.52M parameters and 69.55G FLOPs, compared with 37.09M parameters and 76.57G FLOPs for the baseline. These results confirm that SDLE, DSPC, and EDG-FPN provide complementary improvements and jointly achieve a favorable balance between segmentation accuracy and computational efficiency.

The influence of boundary supervision is further examined by varying the boundary loss coefficient λ_{edge} while keeping the network architecture, class weight vector, Dice loss coefficient, dataset partition, and all other training configurations unchanged. Each candidate setting is trained independently and evaluated on the same fixed OSDD validation subset. When $\lambda_{\text{edge}} = 0$, EdgeHead and its coupling with the BAG are retained, but direct boundary loss supervision is removed. The validation results are presented in Table 4.

Table 4. Sensitivity analysis of the boundary loss coefficient λ_{edge} on the fixed OSDD validation subset. Across the λ_{edge} sweep, the class weights $\mathbf{w} = [0.50, 2.53, 1.46, 3.00, 1.51]$ for sea surface, oil spill, look-alike, ship, and land, respectively, and the Dice loss coefficient $\lambda_{\text{dice}} = 2.0$ remains fixed.

λ_{edge}	Val mIoU (%) ↑	Val Oil IoU (%) ↑
0.0	70.15	56.51
0.1	70.85	58.64
0.2	71.27	60.30
0.4	72.61	60.91
0.6	72.38	60.36
0.8	72.31	60.02

Note: ↑ indicates that higher values are better.

As shown in Table 4, removing direct boundary supervision results in a validation mIoU of 70.15% and a validation Oil IoU of 56.51%. Increasing λ_{edge} from 0 to 0.4 progressively improves both metrics, with the highest validation mIoU of 72.61% and validation Oil IoU of 60.91% obtained at $\lambda_{\text{edge}} = 0.4$. Compared with the setting without direct boundary supervision, this configuration improves the validation mIoU and Oil IoU by

2.46 and 4.40 percentage points, respectively. These results demonstrate that the direct supervision of EdgeHead provides useful boundary cues for improving the delineation of weak and fragmented oil film regions.

Further increasing λ_{edge} to 0.6 or 0.8 does not provide an additional improvement. The corresponding validation mIoU values decrease slightly to 72.38% and 72.31%, while the validation Oil IoU decreases to 60.36% and 60.02%, respectively. This result suggests that an excessively large boundary loss contribution may shift the optimization toward local edge responses at the expense of region-level semantic consistency. Therefore, $\lambda_{\text{edge}} = 0.4$ is selected exclusively according to the validation results and subsequently fixed for final evaluation on the official test split; the test split is not involved in hyperparameter selection.

4. Discussion

The results on the OSDD highlight the value of coordinating feature refinement and decoder design for SAR oil spill segmentation. With the same ConvNeXtV2-Tiny backbone, SDE-Net improves the mIoU and Oil IoU over the UPerNet baseline by 4.25 and 8.91 percentage points, respectively, while using fewer parameters and lower FLOPs (Table 2). The pronounced improvements in oil spill and ship segmentation are particularly relevant for maritime scenes dominated by sea surface pixels, where the accurate representation of sparse foreground targets is important. These findings support task-oriented feature processing as an effective approach to improving the segmentation accuracy within a compact architecture.

For the OSDD comparison, the ConvNeXt-Small encoder of FCS-Net was replaced with ConvNeXtV2-Tiny to match the backbone architecture of SDE-Net. Compared with this adapted implementation, SDE-Net improves the mIoU and Oil IoU by 2.90 and 6.32 percentage points, respectively, while reducing the parameter count by 9.09% and FLOPs by 28.93% (Table 2), indicating a favorable accuracy–complexity trade-off.

The ablation results provide further insight into the complementary contributions of the proposed modules. SDLE improves the segmentation of oil spills and ships through directional local refinement, while DSPC achieves the strongest overall performance among the individual module additions, supporting the value of input-adaptive contextual aggregation (Table 3). EDG-FPN contributes to the segmentation performance while reducing decoder complexity. Their combined configuration achieves the highest mIoU and mDice among the tested variants and improves the Look-Alike IoU relative to SDLE alone. Together, these findings support the combination of the three modules to capture local shapes, use context at different scales, and guide feature fusion with predicted boundaries. Despite the overall gains, the Look-Alike IoU remains 0.83 percentage points below that of the same-backbone UPerNet baseline, indicating room for further improvement in distinguishing among visually similar dark formations.

The boundary supervision results also support the incorporation of structural information into the training objective. Among the tested coefficients, $\lambda_{\text{edge}} = 0.4$ yields the highest validation performance, improving the validation Oil IoU by 4.40 percentage points relative to the setting without direct boundary supervision (Table 4). Because the boundary targets are derived directly from existing segmentation labels, this supervision can be incorporated without additional manual edge annotation. It therefore provides a practical means of complementing region-level optimization with boundary-related information within the proposed framework.

Overall, the combination of improved foreground segmentation and reduced computational requirements relative to the same-backbone baseline provides a useful basis for the further development of SAR oil spill monitoring methods. The present evaluation establishes performance within the OSDD, while transfer across sensors and independent sea

regions remains to be assessed. Cui et al. investigated cross-event and cross-sensor oil spill segmentation using unsupervised adaptation with iterative pseudo-label refinement [60], while HA-CP-Net was used to explore transfer from Sentinel-1 to RADARSAT-2 scenes across sea regions using few labeled target-domain samples [61]. These studies motivate the future assessment of SDE-Net across sensors and sea regions using domain adaptation.

5. Conclusions

In this paper, we propose SDE-Net, a task-oriented semantic segmentation network for multiclass SAR oil spill interpretation under complex maritime conditions. The proposed method was developed to address several persistent challenges in SAR oil spill segmentation, including elongated and fragmented oil slick structures, large variations in contextual scale, weak boundaries, and confusion between oil spills and visually similar look-alike formations. Built upon a ConvNeXtV2-Tiny encoder and a redesigned UPerNet-style decoder, SDE-Net integrates three complementary components: the Strip-Directional Local Enhancement (SDLE) module for directional structural refinement, the Dynamic Scale Pyramid Context (DSPC) module for input-adaptive contextual aggregation, and the Edge-Aware Dynamic Gated FPN (EDG-FPN) decoder for boundary-sensitive cross-level feature fusion.

Extensive experiments conducted on the Oil Spill Detection Dataset demonstrate the effectiveness of the proposed approach. SDE-Net achieves a 72.94% mIoU and 82.88% mDice, outperforming the UPerNet baseline equipped with the same ConvNeXtV2-Tiny backbone by 4.25 and 3.84 percentage points, respectively. The proposed method substantially improves the Oil IoU and increases the Ship IoU by 12.68 percentage points over the same-backbone baseline, while maintaining a competitive ship result and preserving its strong performance on the sea surface and land categories. In addition, SDE-Net achieves a favorable accuracy–complexity trade-off, reducing the computational cost compared with the baseline while preserving a compact model size. The ablation experiments further verify that SDLE, DSPC, and EDG-FPN provide complementary contributions to the final segmentation performance, and the boundary supervision strategy effectively improves the delineation of weak and fragmented oil spill regions.

Despite these improvements, distinguishing oil spills from look-alike dark formations remains challenging because both categories may exhibit highly similar SAR backscatter characteristics. Future work will focus on domain-adaptive learning and lightweight network designs for broader SAR oil spill monitoring.

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

Funding: This research was funded by the Nantong Institute of Technology through the Innovation Experimental Class Project (grant number T2025019C).

Data Availability Statement: The Oil Spill Detection Dataset (OSDD) used in this study was obtained from <https://m4d.itl.gr/oil-spill-detection-dataset/> (accessed on 17 November 2025) [15]. The code and trained model weights used in this study are available from the corresponding author upon request.

Acknowledgments: The authors thank the Nantong Institute of Technology for its support of this research.

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

References

1. Chang, S.E.; Stone, J.; Demes, K.; Piscitelli, M. Consequences of Oil Spills: A Review and Framework for Informing Planning. *Ecol. Soc.* **2014**, *19*, 26. [\[CrossRef\]](#)
2. Beyer, J.; Trannum, H.C.; Bakke, T.; Hodson, P.V.; Collier, T.K. Environmental Effects of the Deepwater Horizon Oil Spill: A Review. *Mar. Pollut. Bull.* **2016**, *110*, 28–51. [\[CrossRef\]](#)
3. Burgherr, P. In-Depth Analysis of Accidental Oil Spills from Tankers in the Context of Global Spill Trends from All Sources. *J. Hazard. Mater.* **2007**, *140*, 245–256. [\[CrossRef\]](#)
4. Andrews, N.; Bennett, N.J.; Le Billon, P.; Green, S.J.; Cisneros-Montemayor, A.M.; Amongin, S.; Gray, N.J.; Sumaila, U.R. Oil, Fisheries and Coastal Communities: A Review of Impacts on the Environment, Livelihoods, Space and Governance. *Energy Res. Soc. Sci.* **2021**, *75*, 102009. [\[CrossRef\]](#)
5. Boufadel, M.; Bracco, A.; Chassignet, E.P.; Chen, S.S.; D’Asaro, E.; Dewar, W.K.; Garcia-Pineda, O.; Justić, D.; Katz, J.; Kourafalou, V.H.; et al. Physical Transport Processes That Affect the Distribution of Oil in the Gulf of Mexico: Observations and Modeling. *Oceanography* **2021**, *34*, 58–75. [\[CrossRef\]](#)
6. Xie, M.; Li, Y.; Zhang, Z.; Fu, Q.; Jiang, H. Remote Sensing of the Oil Spills Caused by Ships: A Review. *Mar. Pollut. Bull.* **2025**, *214*, 117754. [\[CrossRef\]](#)
7. Keramea, P.; Spanoudaki, K.; Zodiatis, G.; Gikas, G.; Sylaios, G. Oil Spill Modeling: A Critical Review on Current Trends, Perspectives, and Challenges. *J. Mar. Sci. Eng.* **2021**, *9*, 181. [\[CrossRef\]](#)
8. Brekke, C.; Solberg, A.H.S. Oil Spill Detection by Satellite Remote Sensing. *Remote Sens. Environ.* **2005**, *95*, 1–13. [\[CrossRef\]](#)
9. Jha, M.N.; Levy, J.; Gao, Y. Advances in Remote Sensing for Oil Spill Disaster Management: State-of-the-Art Sensors Technology for Oil Spill Surveillance. *Sensors* **2008**, *8*, 236–255. [\[CrossRef\]](#)
10. Jafarzadeh, H.; Mahdianpari, M.; Homayouni, S.; Mohammadimanesh, F.; Dabboor, M. Oil Spill Detection from Synthetic Aperture Radar Earth Observations: A Meta-Analysis and Comprehensive Review. *GISci. Remote Sens.* **2021**, *58*, 1022–1051. [\[CrossRef\]](#)
11. Zheng, H.; Zhang, Y.; Khenchaf, A.; Wang, Y.; Ghanmi, H.; Zhao, C. Investigation of EM Backscattering from Slick-Free and Slick-Covered Sea Surfaces Using the SSA-2 and SAR Images. *Remote Sens.* **2018**, *10*, 1931. [\[CrossRef\]](#)
12. Gade, M.; Alpers, W.; Hühnerfuss, H.; Wismann, V.R.; Lange, P.A. On the Reduction of the Radar Backscatter by Oceanic Surface Films: Scatterometer Measurements and Their Theoretical Interpretation. *Remote Sens. Environ.* **1998**, *66*, 52–70. [\[CrossRef\]](#)
13. Migliaccio, M.; Nunziata, F.; Buono, A. SAR Polarimetry for Sea Oil Slick Observation. *Int. J. Remote Sens.* **2015**, *36*, 3243–3273. [\[CrossRef\]](#)
14. Fingas, M.; Brown, C.E. A Review of Oil Spill Remote Sensing. *Sensors* **2018**, *18*, 91. [\[CrossRef\]](#)
15. Krestenitis, M.; Orfanidis, G.; Ioannidis, K.; Avgerinakis, K.; Vrochidis, S.; Kompatsiaris, I. Oil Spill Identification from Satellite Images Using Deep Neural Networks. *Remote Sens.* **2019**, *11*, 1762. [\[CrossRef\]](#)
16. Cantorna, D.; Dafonte, C.; Iglesias, A.; Arcay, B. Oil Spill Segmentation in SAR Images Using Convolutional Neural Networks. A Comparative Analysis with Clustering and Logistic Regression Algorithms. *Appl. Soft Comput.* **2019**, *84*, 105716. [\[CrossRef\]](#)
17. Trujillo-Acatitla, R.; Tuxpan-Vargas, J.; Ovando-Vázquez, C.; Monterrubio-Martínez, E. Marine oil spill detection and segmentation in SAR data with two steps Deep Learning framework. *Mar. Pollut. Bull.* **2024**, *204*, 116549. [\[CrossRef\]](#)
18. Metheniti, V.; Parasyris, A.; Fazzini, N.; Outmani, S.; Correia, M.; Goddard, J.; Alexandrakis, G.; Kozyrakis, G.V.; Vettorello, L.; Keeble, S.; et al. Coastal Crete: A Digital Twin of the Ocean for Oil Spill Identification and Forecasting. In *Proceedings of the OCEANS 2025 Brest, Brest, France, 16–19 June 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–8. [\[CrossRef\]](#)
19. Wei, L.; Hu, Z.; Dong, L.; Zhao, W. A Damage Assessment Model of Oil Spill Accident Combining Historical Data and Satellite Remote Sensing Information: A Case Study in Penglai 19-3 Oil Spill Accident of China. *Mar. Pollut. Bull.* **2015**, *91*, 258–271. [\[CrossRef\]](#)
20. Solberg, A.H.S.; Storvik, G.; Solberg, R.; Volden, E. Automatic Detection of Oil Spills in ERS SAR Images. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1916–1924. [\[CrossRef\]](#)
21. Al-Ruzouq, R.; Gibril, M.B.A.; Shanableh, A.; Kais, A.; Hamed, O.; Al-Mansoori, S.; Khalil, M.A. Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review. *Remote Sens.* **2020**, *12*, 3338. [\[CrossRef\]](#)
22. Song, D.; Zhen, Z.; Wang, B.; Li, X.; Gao, L.; Wang, N.; Xie, T.; Zhang, T. A Novel Marine Oil Spillage Identification Scheme Based on Convolution Neural Network Feature Extraction from Fully Polarimetric SAR Imagery. *IEEE Access* **2020**, *8*, 59801–59820. [\[CrossRef\]](#)
23. Huang, X.; Zhang, B.; Perrie, W.; Lu, Y.; Wang, C. A Novel Deep Learning Method for Marine Oil Spill Detection from Satellite Synthetic Aperture Radar Imagery. *Mar. Pollut. Bull.* **2022**, *179*, 113666. [\[CrossRef\]](#)
24. Naz, S.; Iqbal, M.F.; Mahmood, I.; Allam, M. Marine Oil Spill Detection Using Synthetic Aperture Radar over Indian Ocean. *Mar. Pollut. Bull.* **2021**, *162*, 111921. [\[CrossRef\]](#)
25. Yang, Y.J.; Singha, S.; Goldman, R.; Schütte, F. Dataset of Oil Slicks, Look-Alikes and Remarkable SAR Signatures Obtained from Sentinel-1 Data in the Eastern Mediterranean Sea. *Earth Syst. Sci. Data* **2025**, *17*, 6807–6837. [\[CrossRef\]](#)

26. Fan, Y.; Rui, X.; Zhang, G.; Yu, T.; Xu, X.; Poslad, S. Feature Merged Network for Oil Spill Detection Using SAR Images. *Remote Sens.* **2021**, *13*, 3174. [\[CrossRef\]](#)
27. Basit, A.; Siddique, M.A.; Bashir, S.; Naseer, E.; Sarfraz, M.S. Deep Learning-Based Detection of Oil Spills in Pakistan's Exclusive Economic Zone from January 2017 to December 2023. *Remote Sens.* **2024**, *16*, 2432. [\[CrossRef\]](#)
28. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015*; IEEE: Piscataway, NJ, USA, 2015; pp. 3431–3440. [\[CrossRef\]](#)
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. Zeng, K.; Wang, Y. A Deep Convolutional Neural Network for Oil Spill Detection from Spaceborne SAR Images. *Remote Sens.* **2020**, *12*, 1015. [\[CrossRef\]](#)
32. Zhang, J.; Yang, P.; Ren, X. Detection of Oil Spill in SAR Image Using an Improved DeepLabV3+. *Sensors* **2024**, *24*, 5460. [\[CrossRef\]](#)
33. Wang, D.; Wan, J.; Liu, S.; Chen, Y.; Yasir, M.; Xu, M.; Ren, P. BO-DRNet: An Improved Deep Learning Model for Oil Spill Detection by Polarimetric Features from SAR Images. *Remote Sens.* **2022**, *14*, 264. [\[CrossRef\]](#)
34. Fan, J.; Liu, C. Multitask GANs for Oil Spill Classification and Semantic Segmentation Based on SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2532–2546. [\[CrossRef\]](#)
35. Li, C.; Yang, Y.; Yang, X.; Chu, D.; Cao, W. A Novel Multi-Scale Feature Map Fusion for Oil Spill Detection of SAR Remote Sensing. *Remote Sens.* **2024**, *16*, 1684. [\[CrossRef\]](#)
36. Chen, Y.; Wang, Z. Marine Oil Spill Detection from SAR Images Based on Attention U-Net Model Using Polarimetric and Wind Speed Information. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12315. [\[CrossRef\]](#)
37. Mahmoud, A.S.; Mohamed, S.A.; El-Khoriby, R.A.; AbdelSalam, H.M.; El-Khodary, I.A. Oil Spill Identification Based on Dual Attention UNet Model Using Synthetic Aperture Radar Images. *J. Indian Soc. Remote Sens.* **2023**, *51*, 121–133. [\[CrossRef\]](#)
38. Zhu, Q.; Zhang, Y.; Li, Z.; Yan, X.; Guan, Q.; Zhong, Y.; Zhang, L.; Li, D. Oil Spill Contextual and Boundary-Supervised Detection Network Based on Marine SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5213910. [\[CrossRef\]](#)
39. Wang, S.; Min, B.W.; Gao, D.; Hong, Y. FCS-Net: A Frequency-Spatial Coordinate and Strip-Augmented Network for SAR Oil Spill Segmentation. *J. Mar. Sci. Eng.* **2026**, *14*, 168. [\[CrossRef\]](#)
40. Song, W.; Ma, X.; Song, W. Automatic Detection of Marine Oil Spills from Polarimetric SAR Images Using Deep Convolutional Neural Network Model. *Ecol. Indic.* **2024**, *169*, 112934. [\[CrossRef\]](#)
41. Wu, W.; Wong, M.S.; Yu, X.; Shi, G.; Kwok, C.Y.T.; Zou, K. Compositional Oil Spill Detection Based on Object Detector and Adapted Segment Anything Model from SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4007505. [\[CrossRef\]](#)
42. Giannopoulos, G.; Kremezi, M.; Karathanassi, V.; Andronis, V.; Bliziotis, D.; Kikaki, K.; Oliveira, A.S.; Mütting, A. Two-Stage Oil Spill Detection in SAR Using a Domain-Adapted Segment Anything Model. *Remote Sens.* **2026**, *18*, 1948. [\[CrossRef\]](#)
43. Shaban, M.; Salim, R.; Abu Khalifeh, H.; Khelifi, A.; Shalaby, A.; El-Mashad, S.; Mahmoud, A.; Ghazal, M.; El-Baz, A. A Deep-Learning Framework for the Detection of Oil Spills from SAR Data. *Sensors* **2021**, *21*, 2351. [\[CrossRef\]](#)
44. Chen, Y.T.; Chang, L.; Wang, J.H. Full-Scale Aggregated MobileUNet: An Improved U-Net Architecture for SAR Oil Spill Detection. *Sensors* **2024**, *24*, 3724. [\[CrossRef\]](#)
45. Alpers, W.; Holt, B.; Zeng, K. Oil Spill Detection by Imaging Radars: Challenges and Pitfalls. *Remote Sens. Environ.* **2017**, *201*, 133–147. [\[CrossRef\]](#)
46. Woo, S.; Debnath, S.; Hu, R.; Chen, X.; Liu, Z.; Kweon, I.S.; Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 16133–16142. [\[CrossRef\]](#)
47. Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; Sun, J. Unified Perceptual Parsing for Scene Understanding. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11209, pp. 432–448. [\[CrossRef\]](#)
48. Liu, W.; Lu, H.; Fu, H.; Cao, Z. Learning to Upsample by Learning to Sample. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 6004–6014. [\[CrossRef\]](#)
49. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 5228–5237. [\[CrossRef\]](#)
50. Yuan, Y.; Xie, J.; Chen, X.; Wang, J. SegFix: Model-Agnostic Boundary Refinement for Segmentation. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 12357, pp. 489–506. [\[CrossRef\]](#)

51. Ishikawa, H.; Aoki, Y. Boosting Semantic Segmentation by Conditioning the Backbone with Semantic Boundaries. *Sensors* **2023**, *23*, 6980. [[CrossRef](#)]
52. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 6230–6239. [[CrossRef](#)]
53. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 510–519. [[CrossRef](#)]
54. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 936–944. [[CrossRef](#)]
55. Parasyris, A.; Metheniti, V.; Kampanis, N.; Darmaraki, S. Marine heatwaves in the Mediterranean Sea: A convolutional neural network study for extreme event prediction. *Ocean Sci.* **2025**, *21*, 897–912. [[CrossRef](#)]
56. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In *Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019*.
57. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; Volume 11211, pp. 833–851. [[CrossRef](#)]
58. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. In *Proceedings of the Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 12077–12090.
59. Guo, M.H.; Lu, C.Z.; Hou, Q.; Liu, Z.; Cheng, M.M.; Hu, S.M. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. In *Proceedings of the Advances in Neural Information Processing Systems*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2022; Volume 35, pp. 1140–1156. [[CrossRef](#)]
60. Cui, G.; Fan, J.; Zou, Y. Enhanced unsupervised domain adaptation with iterative pseudo-label refinement for inter-event oil spill segmentation in SAR images. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *139*, 104479. [[CrossRef](#)]
61. Song, D.; Wang, S.; Wang, B.; Chen, W.; Chen, L. HA-CP-Net: A Cross-Domain Few-Shot SAR Oil Spill Detection Network Based on Hybrid Attention and Category Perception. *J. Mar. Sci. Eng.* **2025**, *13*, 1340. [[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.