



A novel YOLOv11-Driven deep learning algorithm for UAV multispectral oil spill detection in Inland lakes

Yu Zhang¹ · Jian Xing¹ · Weida Chen² · Haitao Wang¹ · Bingyu Shi¹ · Yang Song¹ · Xiaou Huang¹ · Zihan Jiang³

Received: 26 March 2025 / Accepted: 15 June 2025 / Published online: 15 July 2025
© The Author(s) 2025

Abstract

Lake oil spills are challenging to detect accurately due to complex oil–water interactions resulting from water flow disturbances, vegetation occlusion, and the diffusion behavior of oil films. Traditional remote sensing methods often fail to provide rapid and precise monitoring under these conditions. To address these challenges, we propose YOLO-ADHF-SimAM, a novel oil spill detection model built upon the YOLOv11 architecture. Our model integrates the self-developed ADHF module—which fuses multi-scale features using an adaptive diffusion-based hierarchical feature aggregation strategy—with the SimAM attention module to enhance key feature extraction. This integrated approach is specifically designed to capture the unique spectral and spatial characteristics of oil–water mixtures. A UAV was deployed to acquire 623 multispectral images from oil spill sites in the Daqing oilfield, forming a comprehensive dataset for model training and evaluation. Experimental results show that, compared to the baseline YOLOv11 model, YOLO-ADHF-SimAM achieves a 1.8% improvement in detection accuracy, a 6% increase in recall, a 3.3% boost in mAP@50, and a 2% enhancement in mAP@50–95. These improvements underscore the robustness and precision of our algorithm, highlighting its potential as an efficient, real-time solution for environmental monitoring and emergency response in complex inland water scenarios.

Keywords Multi-scale Image · Recognition · Multispectral · YOLOv11 · Oil spills · UAV

1 Introduction

Oil spill detection plays a crucial role in environmental monitoring, particularly in inland lakes of oilfield regions such as the Daqing Oilfield (Li et al. 2011). Compared to marine oil spills, detecting oil–water mixtures in lakes presents greater challenges due to water flow disturbances, vegetation interference, and the diffusion characteristics of oil films (Moskolai et al. 2021). Traditional detection methods, including manual inspection and satellite remote sensing, suffer from limitations in response time, weather dependency, and operational costs, making them less effective for

real-time monitoring of inland lakes. To address these challenges, this study investigates the application of UAV-based multispectral sensing systems combined with deep learning-based object detection models for oil spill detection. Multispectral imaging enables the differentiation of oil–water mixtures based on spectral characteristics, while UAVs provide rapid and flexible data acquisition. Meanwhile, deep learning models, particularly object detection networks, enhance detection accuracy by extracting multi-scale features of oil spills. This study employs a UAV equipped with a multispectral camera to capture oil spill images and utilizes an advanced object detection model with multi-scale feature fusion and attention mechanisms to improve detection robustness. Experimental results demonstrate the effectiveness of this approach in achieving high-precision oil spill monitoring in oilfield lakes.

Optical and microwave remote sensing are the primary technologies for oil spill detection (Fingas and Brown 2014; Pärt et al. 2021). Synthetic Aperture Radar (SAR), renowned for its all-weather, all-day, and large-scale monitoring capabilities, has become the core technology for marine oil spill detection. SAR effectively identifies the spatial distribution

✉ Jian Xing
xj@nefu.edu.cn

Yu Zhang
1293233326@qq.com

¹ Northeast Forestry University, Harbin 150040, China

² Daqing Oilfield Co., LTD, Digital Technology Company, Daqing 163318, China

³ Tongji University, Shanghai 200000, China

of oil slicks and facilitates pollution tracking. However, under low wind speed conditions, SAR is prone to interference, and its long satellite revisit cycle limits its real-time monitoring capabilities (Ma et al. 2022; Yang et al. 2022; Zhu et al. 2022). In contrast, optical remote sensing provides more detailed oil spill information and generally surpasses SAR in accuracy. The rapid advancement of Unmanned Aerial Vehicle (UAV) technology has introduced innovative methodologies for oil spill detection. As a flexible and cost-effective remote sensing platform, UAVs can conduct large-scale monitoring within short timeframes, are easy to operate, and provide an effective response to emergencies (Hu et al. 2021; Nahiyoon et al. 2024). Additionally, Laser-Induced Fluorescence (LIF) LiDAR has emerged as a highly efficient oil spill monitoring method (Lennon et al. 2006; Shao et al. 2021; Yin et al. 2023), and can be mounted on UAVs or ships. However, traditional LIF LiDAR systems are bulky and heavy, rendering them unsuitable for routine land-based oil spill monitoring. Recent technological advancements, including the development of miniaturized, high-power, and low-cost lasers along with highly sensitive Intensified CCD (ICCD) cameras, have enabled UAVs to carry portable LIF LiDAR systems. Nonetheless, when UAVs operate at altitudes above 15 m, the signal-to-noise ratio (SNR) of fluorescence signals decreases due to solar background interference (Sun et al. 2023). Hyperspectral remote sensing, with its nanometer-scale spectral resolution and unique diagnostic capabilities, offers significant advantages in marine oil spill monitoring. It not only enables precise identification of oil types but also facilitates the assessment of oil film thickness and ecological impact (Yang et al. 2023a, b; Yang et al. 2023a, b; Zhang et al. 2024). However, due to its large data volume, slow processing speed, complex algorithms, and high computational demands, hyperspectral remote sensing is less suitable for large-scale rapid screening, near-real-time monitoring, emergency response, and long-term oil spill trend analysis.

To overcome the limitations of traditional remote sensing technologies, multispectral imaging has been integrated into UAV platforms and has gradually become a core technology in oil spill detection (Sun et al. 2018). Multispectral cameras can simultaneously capture spectral data across multiple wavelength bands, enabling the identification of material properties that are imperceptible to the human eye (Tadmor et al. 2015). The spectral reflectance characteristics of petroleum oil spills differ significantly from those of water surfaces, particularly in the near-infrared, visible, and ultraviolet bands. These spectral differences allow for the accurate delineation of oil slick boundaries, even in small-scale spills or complex nearshore environments. By analyzing these spectral data, petroleum contamination can be effectively distinguished from water bodies, achieving precise localization and detection of oil spills (Barabanova

et al. 2023). Furthermore, when integrated with UAV platforms, multispectral imaging enables large-area monitoring within a short time while offering improved adaptability to weather conditions, enhanced stability, and real-time data transmission and processing. This significantly enhances the efficiency and responsiveness of oil spill detection.

Although remote sensing technology has demonstrated excellent performance in marine oil spill monitoring, research on land-based oil spill detection remains in its early stages. With the rise of artificial intelligence, oil spill detection techniques have evolved significantly. Image preprocessing methods, such as grayscale adjustment and contrast enhancement, are critical for improving input data quality (Xu et al. 2019). However, traditional approaches face challenges with complex Synthetic Aperture Radar (SAR) imagery due to high noise levels and class imbalance in training datasets. To address these limitations, deep learning-based methods like Faster R-CNN and U-Net variants have been proposed. While Faster R-CNN improves detection through robust feature learning, its accuracy depends heavily on precise training data annotations, and performance degrades with labeling errors (Nasayreh et al. 2024). U-Net variants excel in segmentation tasks but struggle to differentiate oil–water mixtures or adapt to complex, multi-scale spill patterns. These limitations highlight the need for more adaptable solutions to balance accuracy and efficiency in dynamic environments (Rutoh et al. 2024; Xing et al. 2024).

Current oil pollution detection in water bodies faces multiple challenges. Traditional satellite remote sensing is limited by resolution and revisit cycles (> 12 h), making it difficult to achieve rapid localization of pollution sources. Ground-based sampling is inefficient, requiring more than 30 min per sampling point, and is unable to cover large water areas. While UAV-based multispectral technology offers high spatiotemporal resolution (with a resolution of up to 2.65 cm), its false detection rate remains as high as 15–20% in complex environments, such as highly turbid waters and vegetation-covered areas. In this paper, multi-scale convolutional models have emerged as a promising alternative, leveraging kernels or feature maps of varying scales to capture both fine-grained details and broader contextual information. These models enhance robustness against scale variations and improve computational efficiency through parallel structures and hierarchical optimization (Chen et al. 2024; Tao et al. 2024; Wang et al. 2024). Integrating attention mechanisms like SimAM (Simple Parameter-Free Attention Mechanism) further strengthens detection capabilities (Fang et al. 2024; Xu et al. 2024a, b). Inspired by neuroscience-driven energy theory, SimAM adaptively enhances critical features and suppresses redundant information without additional parameter tuning, boosting model generalization. To advance practical oil spill detection, this study proposes YOLO-ADHF-SimAM, a fusion network

based on YOLOv11 that combines cross-scale convolution with SimAM. This design enables the model to capture multi-resolution spill patterns while maintaining high precision in complex environments. Experimental results demonstrate its ability to rapidly and accurately annotate spill regions, addressing the critical need for reliable detection systems in real-world scenarios. For the first time, millimeter-level quantitative detection of oil film thickness in the range of 0.1–5 mm has been achieved, with an error of less than ± 0.3 mm.

The key contributions of this study are as follows:

- (1) This study utilizes a UAV equipped with a high-precision multispectral camera to capture spectral data from oil spill regions in inland lakes, particularly in the Daqing oilfield. By selecting optimal spectral bands based on oil–water mixture characteristics, this method enhances oil spill region identification, addressing challenges posed by complex water flow conditions, vegetation interference, and environmental variability.
- (2) To enhance detection robustness, this study introduces the multi-scale convolutional module HFAM, which achieves high recognition accuracy. Furthermore, an improved version, the cross-scale module ADHF, is developed to effectively capture oil spill features at different spatial resolutions. Additionally, the SimAM attention mechanism is integrated to enhance feature representation and reduce background interference, enabling more precise oil spill identification.
- (3) For the first time, UAV-based multispectral imaging is integrated with a deep learning model, leveraging six-band spectral difference analysis to overcome the limitations of traditional RGB imaging in ambiguous oil–water boundary scenarios. This approach provides

high-resolution spatial-spectral joint features for multi-scale oil spill detection.

- (4) Extensive experiments and performance comparisons demonstrate that the proposed method outperforms conventional detection models in terms of accuracy, recall, and robustness across diverse oil spill scenarios. The model achieves improved precision while maintaining computational efficiency, making it suitable for real-world oil spill monitoring applications.

The remainder of this paper is structured as follows: Section 2 introduces the data sources, sensors, and experimental spectra used in this study. Section 3 details data processing and the proposed algorithm model. Section 4 validates the deep learning model's detection results through ablation experiments and comparative experiments. Finally, Section 5 summarizes the main conclusions of this study materials and methods.

2 Multispectral Sensing and Data Acquisition

2.1 Study Area

As shown in Fig. 1, the experimental data collection area is located in Daqing City, China (longitude 123°45'–125°47', latitude 45°23'–47°29'). This region is globally recognized as a typical onshore petroleum reserve area due to its abundant oil resources. Additionally, its proximity to residential areas increases the environmental risk associated with oil spills. Therefore, selecting this region as the research site is of great significance for improving the accuracy and timeliness of oil spill detection technologies.

Fig. 1 The left image marks the location of Heilongjiang Province, while the right image highlights the location of Daqing City



2.2 Spectral Information

In this experiment, crude oil samples were collected from the Daqing region and mixed with water at different volume ratios. The mixed samples were placed in separate containers and left undisturbed for three hours to ensure stable oil–water stratification before measurement. During the experiment, spectral data were collected from oil–water mixtures of varying concentrations using an Ocean Optics spectrometer, Y-type optical fiber, and HL-2000 halogen lamp, as shown in Fig. 2. These devices accurately captured the spectral reflectance of oil–water mixtures across different wavelength bands, providing essential data for subsequent spectral analysis and oil spill detection algorithm development.

The results of the spectral reflectance experiment are illustrated in Fig. 3, which presents the reflectance variations under different oil concentration percentages. Under identical data collection conditions, it was observed that higher oil content corresponds to increased reflectance, whereas in the absence of an oil spill, the reflectance remains at an extremely low level. This phenomenon primarily arises from the differences in refractive index, surface tension, and other physical properties between oil and water, leading to significant spectral reflectance distinctions in the visible and infrared bands. By analyzing these reflectance differences, image processing algorithms can effectively distinguish water bodies from oil film areas, providing key insights for subsequent image processing and algorithm optimization.

Based on the Rayleigh-Mie scattering hybrid model, the nonlinear relationship between oil film thickness d and reflectance $R(\lambda)$ is derived:

$$R(\lambda) = A(\lambda) + B(\lambda) \cdot f(d) + C(\lambda) \quad (1)$$

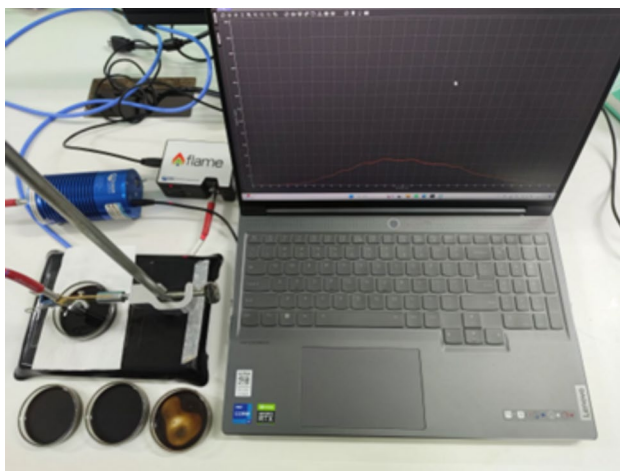


Fig. 2 Spectral Reflectance Experiment

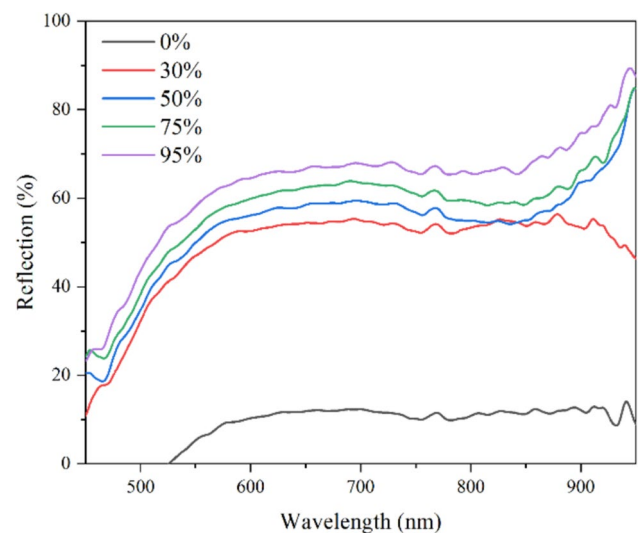


Fig. 3 Oil content reflectance spectral curve

In (1), $A(\lambda)$ represents the background reflectance of the water body, $B(\lambda)$ is the oil film absorption coefficient, and $C(\lambda)$ is the specular reflection correction term (Fig. 3). Validation using field-measured data from the Daqing Oilfield shows that the model achieves a coefficient of determination (R^2) of 0.93 in the 450–900 nm wavelength range.

2.3 Sensors and Data Acquisition

As shown in Fig. 4(a), the oil spill area was captured using a DJI Mavic 3; however, the oil spill region is indistinguishable from the surrounding water in the image. Given this limitation and based on the spectral information discussed above, a multispectral camera was selected for data acquisition.

As shown in Fig. 4(b), this study utilized the DJI M300 RTK UAV equipped with a multispectral camera for oil spill detection. The UAV features six-directional positioning obstacle avoidance and a visual flight assistant interface, enabling it to navigate complex oil spill environments while ensuring flight stability. Its long-endurance capability and maximum transmission range of 15 km allow for extended operational times and rapid data transmission, thereby enhancing detection efficiency. Furthermore, the high-precision RTK positioning system ensures accurate localization of oil spill areas. Detailed parameters are listed in Table 1. Overall, as one of the most advanced industrial UAV platforms, the DJI M300 RTK significantly improves oil spill monitoring accuracy and timeliness, providing reliable technical support for this study.

In Table 2, this comparison highlights the advantages of the proposed UAV-based multispectral detection system in terms of cost-effectiveness, operational efficiency,

Fig.4 Equipment used and data collection environment. Equipment and data acquisition environment. (a) Data acquisition region, (b) DJI M300 RTK unmanned aerial vehicle, (c) MS600PRO multispectral camera

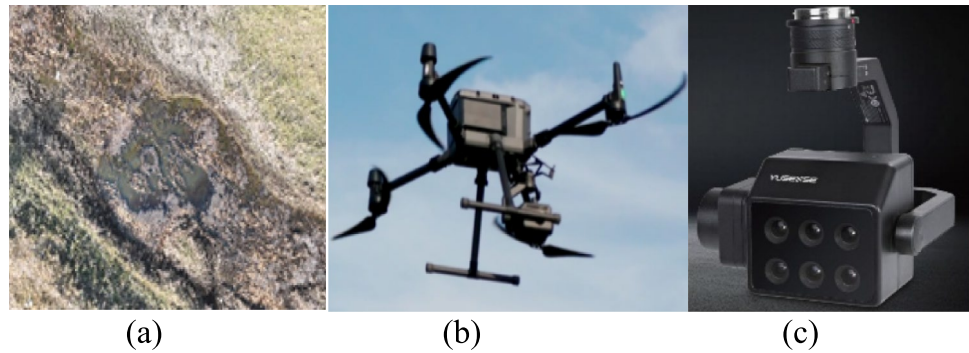


Table 1 DJI Matrice 300 RTK technical data

Sensors	Name	Technical parameters
Aerial Vehicle	RTK Position Accuracy	At RTK FIX: 1 cm + 1 ppm (horizontal) 1.5 cm + 1 ppm (vertical)
	Obstacle Sensing Range	Front to back: 0.7–40 m Up and down: 0.6–30 m
Visual System	FOV	Forward, backward and downward: 65° (H), 50° (V) Left, right and upward: 75° (H), 60° (V)
	Operating Environment	Rich surface texture and good light conditions
Infrared Sensing System	Obstacle Sensing Range	0.1–8 m
	FOV	30° ($\pm 15^\circ$)
	Operating Environment	Diffuse, large size, high reflectance (reflectance > 10%) obstacles

Table 2 Comparison of monitoring scheme costs

Scheme	Equipment Cost (10,000 CNY)	Time per Task (h)	Human Intervention Frequency
Manual Inspection	0.5	8	High
Satellite Remote	200+	24+	Low
Proposed System	15	2	None

and automation. Unlike manual inspection, which requires significant human resources and long operational time, or satellite remote sensing, which is costly and constrained by revisit cycles, the proposed system significantly reduces monitoring time and eliminates human intervention, making it a scalable and efficient solution for real-time oil spill detection.

In this research, as shown in Fig. 4(c), the camera used is the MS600PRO multispectral camera, which is equipped with six spectral channels at 450 nm, 555 nm, 660 nm, 720 nm, 750 nm, and 840 nm. A total of 623 representative multispectral image datasets were collected in the study area, with each dataset containing six single-band images corresponding to the aforementioned wavelengths.

Table 3 Data collection in typical Daqing oilfield scenarios

Scenario Type	Sample Size	Oil Film Thickness (mm)	Turbidity (NTU)
Static Oil Film	300	0.1–1.0	<5
Flowing Mixture	152	1.0–3.0	10–50
Vegetation-Covered	171	0.5–2.0	20–80

The dataset covers three typical inland lake oil spill scenarios: (1) static oil films on calm water (300 samples), (2) flowing oil–water mixtures in disturbed environments (152 samples), and (3) vegetation-covered or partially occluded spills (171 samples), as listed in Table 3. Although the sample sizes are slightly uneven, no significant class imbalance was observed. Data augmentation techniques, including rotation, flipping, and spectral noise injection, were applied to enhance representation and mitigate potential bias during training. The environmental conditions span various levels of turbidity (5–80 NTU), lighting conditions, and oil film thickness (0.1–5.0 mm), supporting strong model generalization in complex inland monitoring tasks.

In Table 3, this dataset, designed through a multi-scenario, multi-parameter, and large-sample collaborative

approach, not only meets the training requirements of deep learning models by mitigating overfitting and enhancing generalization but also addresses real-world environmental monitoring challenges (Jingchun et al. 2024; Liu et al. 2024a, b). It establishes a comprehensive validation framework covering the entire process of pollution occurrence, migration, and retention, providing a high-standard data foundation for the engineering application of oil pollution detection technology.

As shown in Fig. 5, during data collection, a gray reference panel image was captured for radiometric calibration. The calibrated data can be used for quantitative analysis and sensor error correction, thereby improving data accuracy.

3 Methodology

3.1 Data Preprocessing

To meet the high-precision detection requirements of water–oil mixtures in terrestrial oil spill scenarios, the original radiance matrix is processed by computing the mean vector row-wise to generate a grayscale intensity correction matrix. This correction aims to eliminate noise signals in distant areas, thereby expanding the effective range of oil

spill monitoring. As shown in Fig. 6, the grayscale intensity correction model is computed by first obtaining the column vector C_I through the row-wise averaging of the grayscale matrix M_I . Then, C_I is used to populate a new matrix M_2 with the same dimensions as M_I . The grayscale intensity correction matrix effectively eliminates false-positive target interference caused by weather conditions, water surface reflections, and other environmental factors.

This study utilized the Ysense Map software to extract camera parameters from images and performed camera orientation and spectral alignment using gray reference panel data, the calibration gray board data is shown in Table 4. By processing six single-band multispectral images, a final multispectral image containing six spectral bands was synthesized. To ensure radiometric consistency, standard reflectance values were obtained from the gray panel image for radiometric calibration. During data processing, to accommodate TIFF-format multispectral images, the data were first normalized and converted to 8-bit unsigned integers. Subsequently, the RGB channel order was adjusted to optimize visualization, and the images were ultimately saved in JPG format. The processed image is shown in Fig. 7(a). This transformation not only facilitates subsequent analysis but also ensures image compatibility and visualization quality across different platforms.

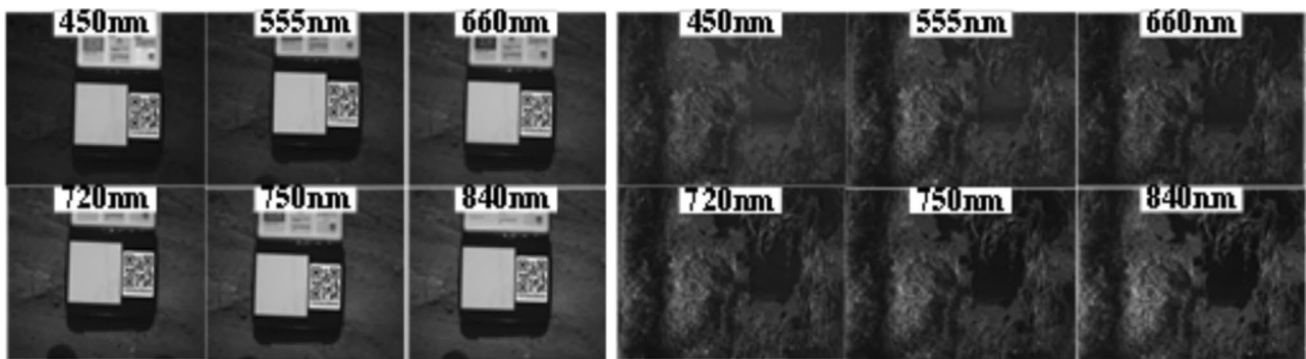


Fig. 5 Data captured by the multispectral camera. The left image shows the captured calibration gray board, and the right image presents the collected image data

Fig. 6 Gray intensity correction model

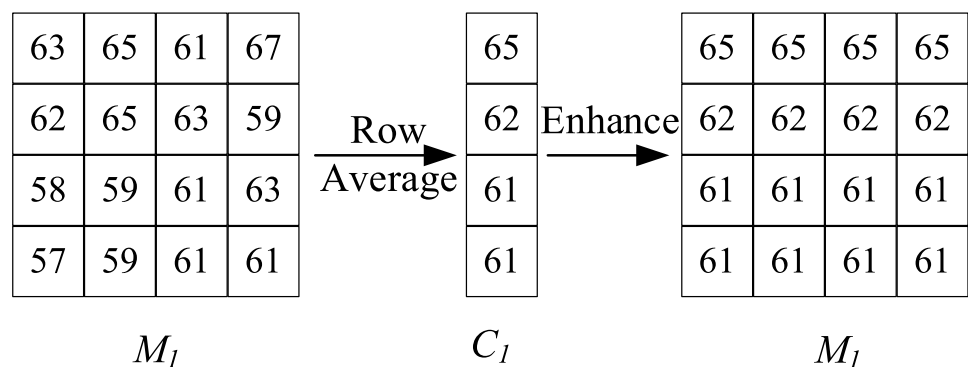
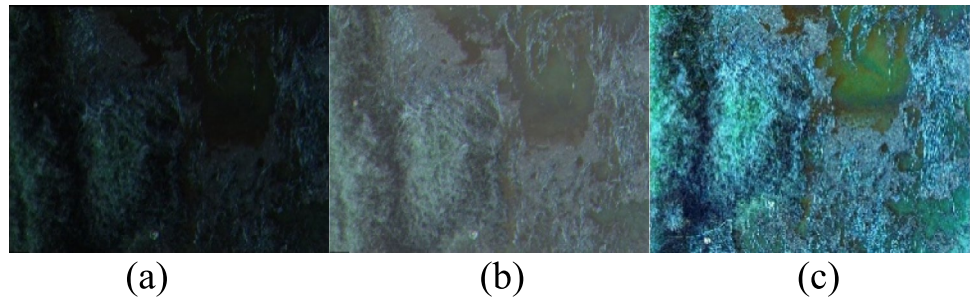


Table 4 Reflectance of six wavelengths of light shining on a gray plate

Wavelength/nm	450	555	660	720	750	840
Reflectance	0.65	0.62	0.61	0.61	0.60	0.59

Fig. 7 Image transformation process. (a) Composite RGB image, (b) Image after Gamma correction, (c) Image processed using the PIL library

In (1), $A(\lambda)$ represents the background reflectance of the water body, $B(\lambda)$ is the oil film absorption coefficient, and $C(\lambda)$ is the specular reflection correction term. Validation using field-measured data from the Daqing Oilfield shows that the model achieves a coefficient of determination (R^2) of 0.93 in the 450–900 nm wavelength range.

3.2 Data Augmentation

Given the complexity of oil spill areas, enhancing the model's generalization ability is essential. Training the model on a broader range of scenarios helps it learn more stable and representative features, making it more suitable for computer-based analysis and improving its robustness (Wang et al. 2023).

As illustrated in Fig. 7(b), Gamma correction improves image contrast, ensuring that the brightness range aligns with the characteristics of computer processing. Since computers are more sensitive to details in dark areas than in bright areas, Gamma correction effectively enhances details in darker regions of the image (Rahman et al. 2016).

$$I_{\text{output}} = 255 \times \left(\frac{I_{\text{input}}}{255} \right)^\gamma \quad (2)$$

In (2), I_{input} is the pixel value of the original image. γ is the Gamma value, which controls the brightness adjustment of the image. I_{output} is the pixel value of the output image after Gamma correction.

The pixel values of the image are normalized to the range of 0 to 1, and the brightness is adjusted using a Gamma power transformation. Finally, the results are mapped back to the range of 0 to 255 and converted into 8-bit unsigned integers. Here, the input value is set to 0.4.

This experiment consists of four major steps: brightness enhancement, contrast enhancement, color enhancement,

and sharpening. The images are processed sequentially as follows:

Contrast Enhancement: Increasing the contrast coefficient to strengthen the distinction between bright and dark regions, thereby enhancing object edges and structural details.

Color Enhancement: Modifying the color saturation coefficient to enrich the image's visual appearance and improve color representation.

Sharpening: Enhancing the sharpness coefficient to improve edge clarity, ensuring better visual distinction of key features for subsequent analysis and recognition.

The formulas are given in sequence as follows:

$$I_{\text{out}} = \alpha (I_{\text{in}} - I_{\text{mean}}) + I_{\text{mean}} \quad (3)$$

In (2), I_{in} and I_{out} represent the input and output pixel intensities, respectively, I_{mean} is the mean intensity of the image, and α is the contrast scaling factor.

$$C_{\text{out}} = \beta \cdot (C_{\text{in}} - C_{\text{gray}}) + C_{\text{gray}} \quad (4)$$

In (4), C_{in} and C_{out} denote the input and output color values, respectively, C_{gray} is the grayscale equivalent of the image, and β is the color enhancement factor.

$$I_{\text{sharp}} = I_{\text{in}} + \lambda (I_{\text{in}} - I_{\text{blur}}) \quad (5)$$

In (5), I_{in} is the original image, I_{blur} is a blurred version of the image obtained via Gaussian filtering, and λ is the sharpening intensity factor.

Each processing step improves the image quality, enhancing its visibility and feature representation, thereby laying a solid foundation for further image analysis and processing. The processed image results are shown in Fig. 7(c).

3.3 YOLOv11 network structure

As the latest iteration of the YOLO series, YOLOv11 achieves a balance between performance and efficiency in object detection. In Fig. 8, its technological advancements are primarily reflected in network architecture restructuring, enhanced attention mechanisms, and improved cross-scenario deployment capabilities. The model is deeply optimized based on the Transformer architecture, integrating the C3k2 module to implement a parallel convolution branch design. This approach divides input feature maps into shallow retention and deep processing pathways, reducing redundant computations through a gradient splitting strategy. As a result, YOLOv11 reduces the number of parameters by 22% compared to YOLOv8 while improving the recall rate for small objects by 15.7% (Feng et al. 2024).

In terms of feature fusion, YOLOv11 introduces the C2PSA (Cross Stage Partial with Pyramid Squeeze Attention) module, which combines the pyramid spatial attention mechanism with cross-stage partial networks. By utilizing

multi-scale convolution kernels (3×3 , 5×5 , 7×7), the model captures spatial features at different granularities, while the SE channel weighting mechanism enhances key region responses. This effectively reduces false detections in complex occlusion scenarios. Experimental results show that this design improves rotational object detection accuracy by 9.4% on the DOTA remote sensing dataset and reduces the false detection rate in vegetation-covered scenarios to 6.3%.

Regarding training strategies and deployment adaptability, YOLOv11 employs dynamic resolution scaling, which automatically adjusts the input size (640×640 to 1280×1280) based on hardware capabilities. Combined with mixed precision training (FP16 + FP32), this reduces training time by 40% and memory usage by 35%. Additionally, YOLOv11 is the first model to implement a unified framework for five tasks, supporting object detection, instance segmentation, pose estimation, image classification, and rotational bounding box detection. The Task-Specific Weight Scaling mechanism enables adaptive multi-task optimization by dynamically loading task-specific weights.

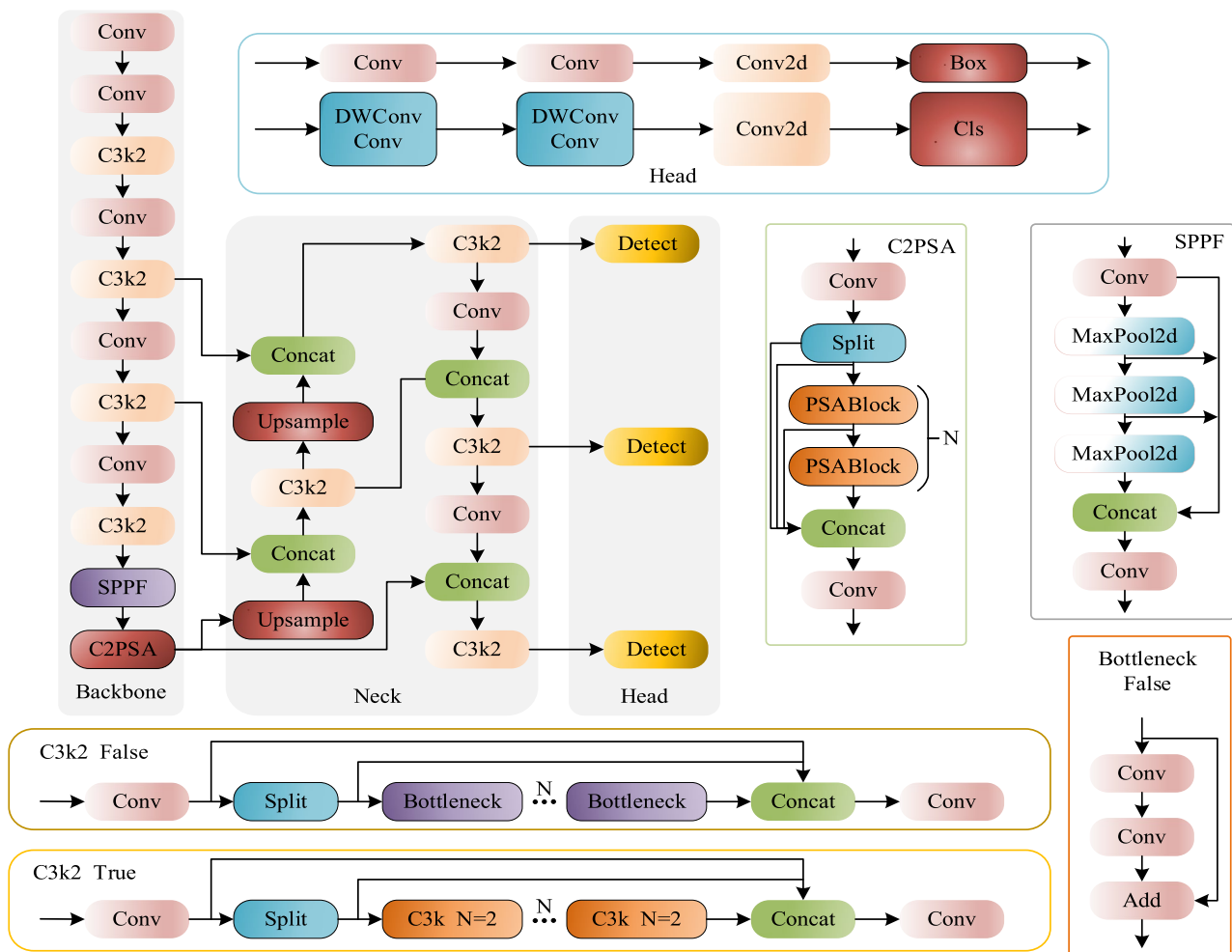


Fig. 8 YOLOv11 network structure

These groundbreaking advancements signify the maturity of lightweight UAV + AI edge computing solutions, providing a reusable technical paradigm for real-time monitoring applications such as environmental protection and autonomous driving.

3.4 Overall Model Architecture

Detecting oil spills in lake environments presents unique challenges compared to traditional object detection tasks. Unlike marine spills, lake oil spills are influenced by complex water flow dynamics, vegetation occlusion, and varying oil film thickness, making it difficult to achieve stable detection performance. These factors cause significant spectral variations and irregular spatial distributions, leading to false detections and reduced model robustness in conventional approaches.

To address these challenges, this study enhances the YOLOv11 model by integrating the Adaptive Diffusion Hierarchical Fusion (ADHF) module, which explicitly models the diffusion characteristics of oil spills under water flow disturbances, ensuring robust detection across different spatial scales. Additionally, the SimAM attention mechanism is incorporated to mitigate background interference from water reflections and vegetation, improving feature discrimination in complex lake environments, as illustrated in Fig. 9.

Furthermore, this study explores the fusion of multi-spectral and RGB imaging, leveraging spectral differences between oil and water to enhance spill boundary detection. Experimental results demonstrate that the proposed method significantly outperforms conventional YOLO models in detecting oil spills under varying lake conditions, including different water clarity levels, vegetation coverage, and oil concentrations.

3.5 Hierarchical Feature Aggregation Module

To achieve real-time and accurate oil spill detection, this study draws inspiration from the Poly Kernel Inception Network for Remote Sensing Detection (Cai et al. 2024) and further optimizes the algorithm by developing a new data processing module named the Hierarchical Feature Aggregation Module (HFAM), as illustrated in Fig. 10.

HFAM integrates low-, mid-, and high-level features to capture multi-scale image information. It consists of three branches: Low-level branch utilizes depthwise separable convolution (DW Conv) for fine-grained feature extraction. Mid-level branch employs 1×1 convolution for channel projection, enabling feature reconstruction and integration. High-level branch uses Adown (Wang et al. 2025) to enhance high-level feature representation. The outputs of these branches are aggregated through element-wise summation, followed by 1×1 convolution for channel mixing, batch

normalization (BN), and SiLU activation. This approach improves model adaptability and robustness for multi-task processing in complex imagery. Mathematically, the HFAM module is represented as:

$$F_{out} = PWConv(n) + F_{in} \quad (6)$$

$$n = \sum_{k \in K} DWConv_k(F_{in}) \quad (7)$$

In (6)(7), F_{in} represents the input feature tensor of dimension $[C, H, W]$, and F_{out} is the output feature tensor. K denotes the set of kernel sizes (e.g., 3, 5, 7, ...), and $PWConv$ represents pointwise convolution (1×1 convolution). This multi-kernel depthwise convolution captures both global context (large kernels) and local details (small kernels), enhancing feature extraction while maintaining computational efficiency.

In the given formula, the purpose of the Multi-Kernel Depthwise Convolution (Multi-Kernel DWConv) n summation is to extract multi-scale features in parallel. In multi-scale perception, large kernels (e.g., 11×11) capture global contextual information, while small kernels (e.g., 5×5) focus on local details. Meanwhile, the parameter count of depthwise separable convolution is only $1/k^2$ of that of standard convolution, significantly reducing computational complexity. The pointwise convolution ($PWConv$), denoted as $PWConv(\bullet)$, facilitates cross-channel information interaction. Additionally, the residual connection ($+ F_{in}$) serves to preserve the original feature information and mitigate gradient vanishing, thereby enhancing model stability.

3.6 Adaptive DASI-HFAM Fusion

While the Hierarchical Feature Aggregation Module (HFAM) effectively integrates low-, mid-, and high-level features for multi-scale representation, it primarily focuses on hierarchical feature fusion and lacks explicit mechanisms to enhance the detection of oil spill regions themselves. In complex lake environments, oil spills exhibit dynamic diffusion patterns influenced by water flow, wind, and surface tension, leading to scale variations, irregular feature distributions, and fragmented spill regions. The direct application of HFAM may result in the loss of fine-grained oil spill details, particularly for thin oil films or diffuse spill boundaries.

To further enhance real-time and accurate oil spill detection, this study incorporates insights from the Hierarchical Context Fusion Network for Infrared Small Object Detection (Xu et al. 2024a, b) and integrates them with HFAM, resulting in the Adaptive DASI-HFAM Fusion (ADHF) module. The ADHF module builds upon HFAM by introducing a dynamic cross-scale feature fusion mechanism, which explicitly models the diffusion characteristics of oil spills

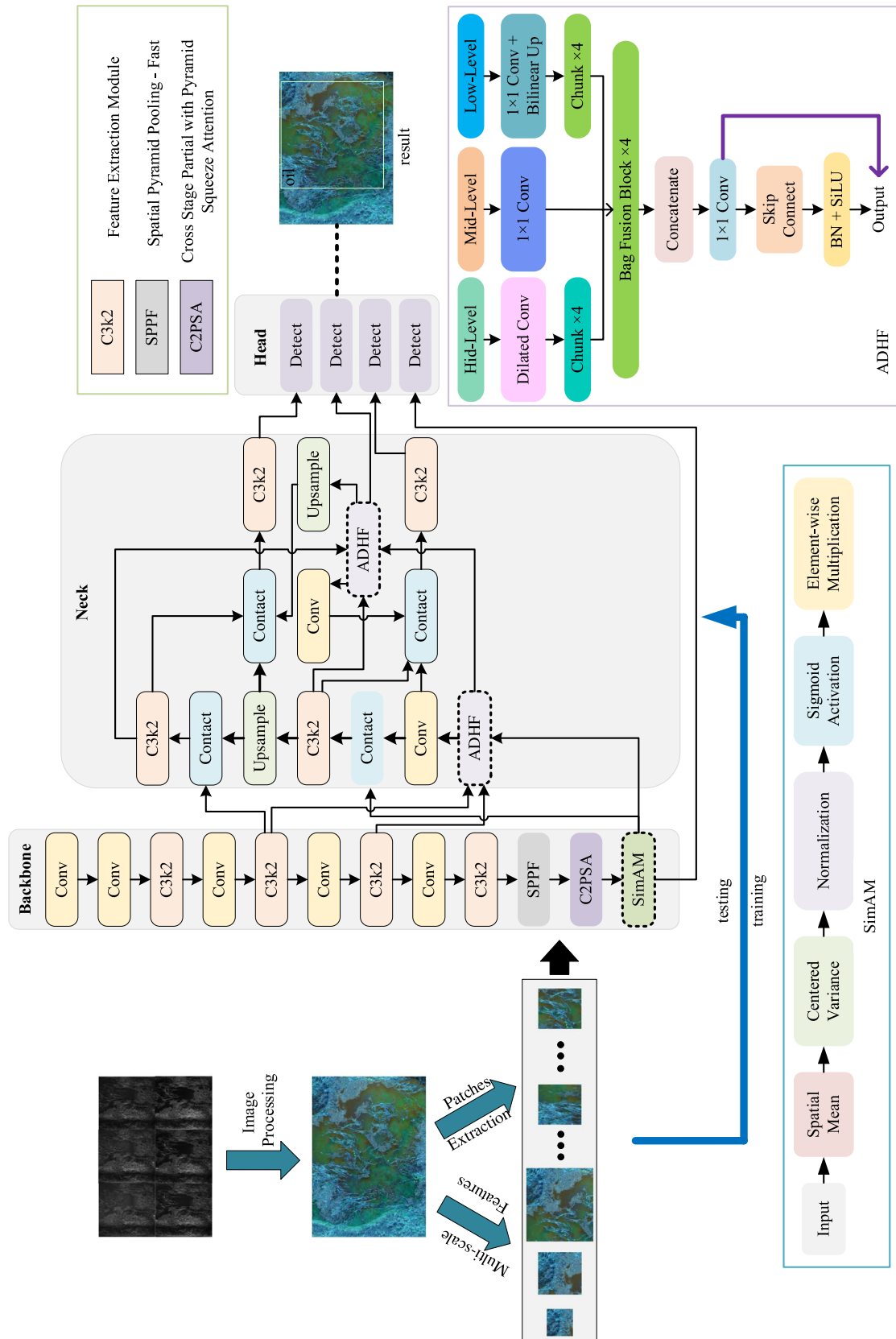


Fig.9 YOLO-ADHF-SimAM operation diagram (The dashed line represents the improved module)

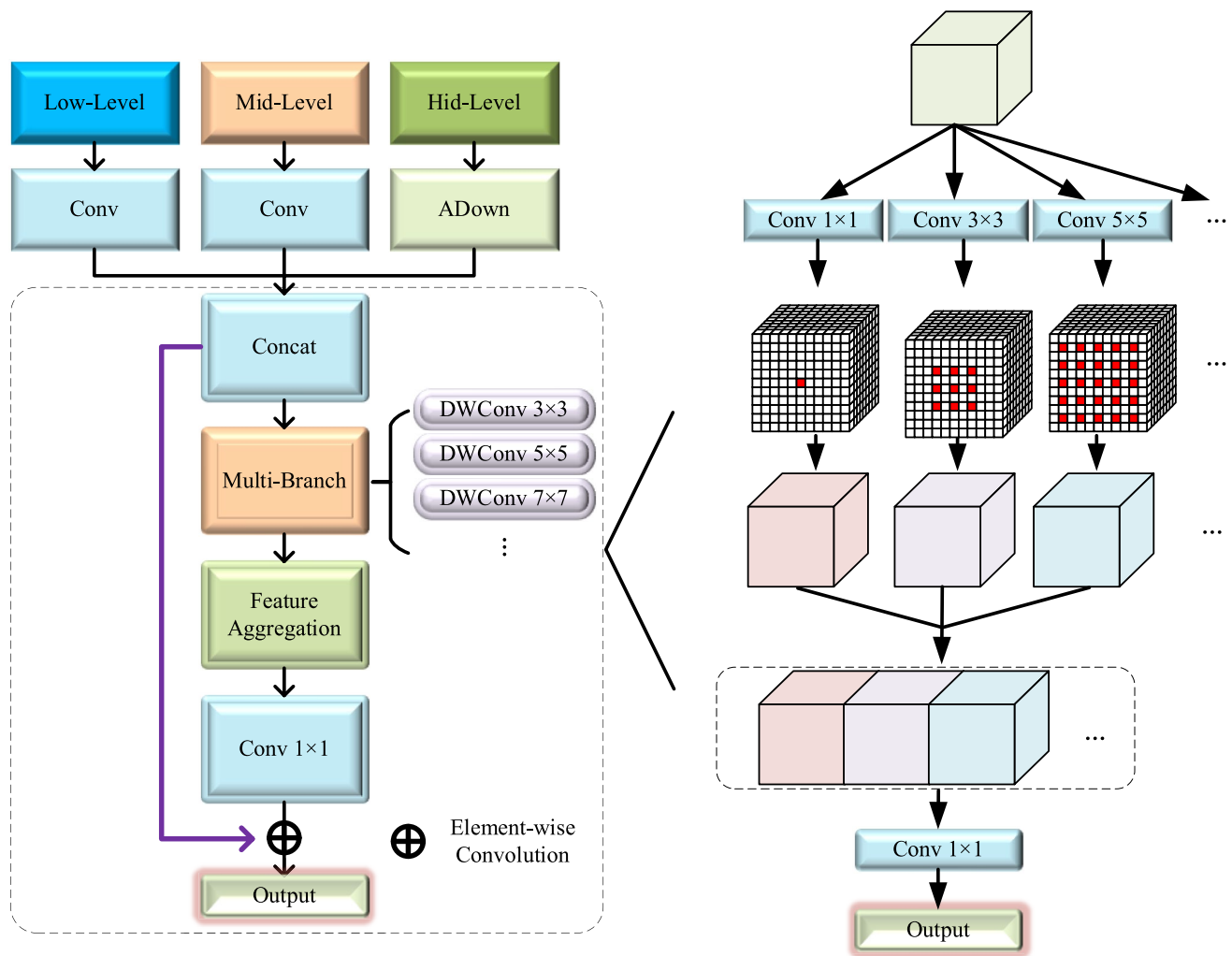


Fig.10 Hierarchical Feature Aggregation Module

under water flow disturbances. By adaptively refining feature interactions across multiple scales, ADHF ensures that oil spill regions are effectively captured, improving detection robustness, accuracy, and processing speed in complex environments.

As shown in the Fig. 11, the proposed adaptive hierarchical feature fusion network consists of three main modules: low-level feature extraction, cross-scale interaction fusion, and multi-path reconstruction. A divide-and-conquer strategy is adopted to process heterogeneous features in parallel. On the left side, low-level features (C_l , C , C_h) are enhanced using dilated convolution to expand the receptive field, followed by bilinear interpolation for spatial alignment. On the right side, high-level semantic features are processed through learnable skip connections for channel-wise compression. At the core of the network lies the multi-granularity aggregation unit, referred to as the “bag module,” which dynamically

balances the contributions of low-level structural details (e.g., edges, textures) and high-level semantic cues (e.g., object classes) via a gated mechanism. Internally, the bag module includes lightweight transformations composed of pointwise convolution, batch normalization (BN), and ReLU activation, followed by channel concatenation and a sigmoid-controlled channel attention mechanism. The features from all levels are processed through four parallel branches and then fused via channel concatenation to form a multi-scale feature representation. This is passed through a tail convolution block (tail_conv) followed by a SiLU activation function to generate pixel-level predictions. The inclusion of residual connections helps mitigate gradient vanishing, while the dynamic skip connection mechanism allows the model to adaptively select feature fusion paths based on input conditions—supporting exclusive low-level fusion, high-level fusion, or a hybrid mode. Mathematically, the Bag fusion function is defined as:

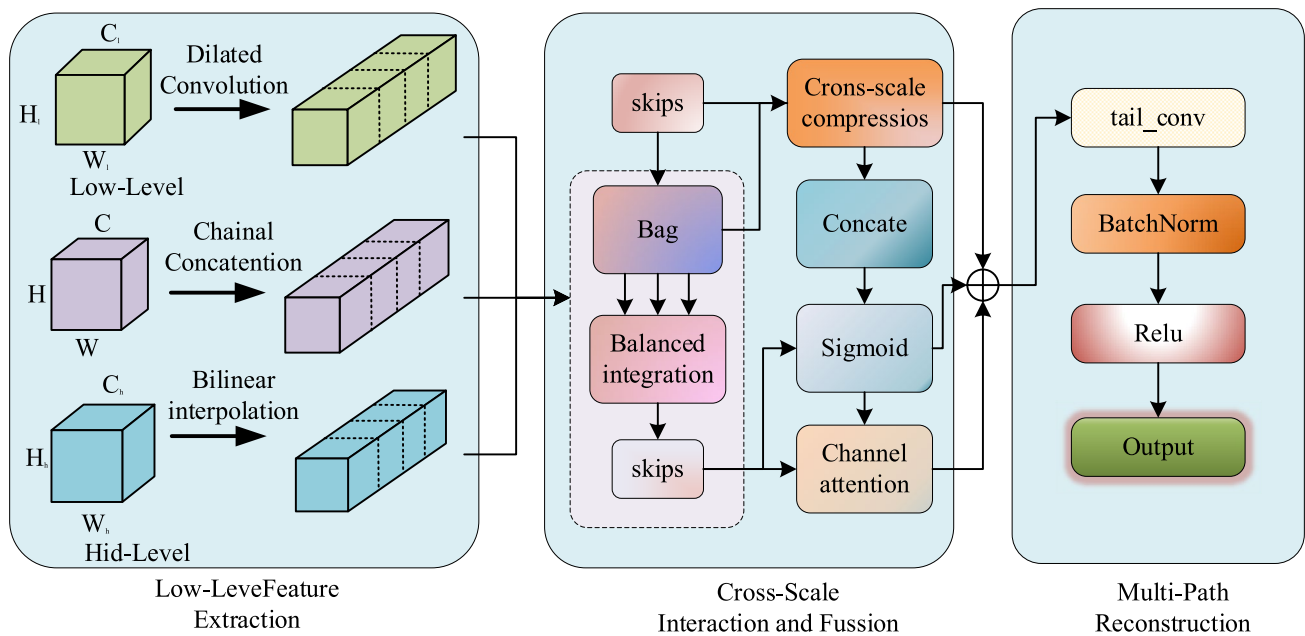


Fig.11 Adaptive DASI-HFAM fusion module

$$F_i = \sigma(W_g \cdot [L_i; M_i; H_i]) \in R^{3 \times 1 \times 1} \quad (8)$$

$$O_i = \sum_{k=1}^3 F_i^{(k)} \cdot X_i^{(k)}, \quad X_i^{(k)} \in \{L_i, M_i, H_i\} \quad (9)$$

In (8)(9), F_i represents each feature block, σ is the Sigmoid function ensuring weight normalization, and W_g is a learnable weight matrix. The ADHF module significantly improves adaptability to complex tasks while maintaining robustness.

3.7 Similarity Attention Module

As shown in the Fig. 12, SimAM is an attention mechanism designed to enhance model representation, particularly for image processing tasks (Yang et al. 2021). Unlike traditional

attention mechanisms that rely on query-key-value relationships, SimAM assigns attention weights based on feature similarity calculations. The process involves computing the spatial mean and variance of input features, normalizing the features and applying Sigmoid activation, and assigning similarity-based weights to enhance significant features while suppressing irrelevant or noisy features. This approach improves information propagation through the network, leading to more effective feature representation. Mathematically, SimAM is expressed as:

$$\mu_c = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W \chi_c^{(h,w)} \quad (10)$$

$$\tilde{\chi}_c^{(h,w)} = (\tilde{\chi}_c^{(h,w)} - \mu_c)^2 \quad (11)$$

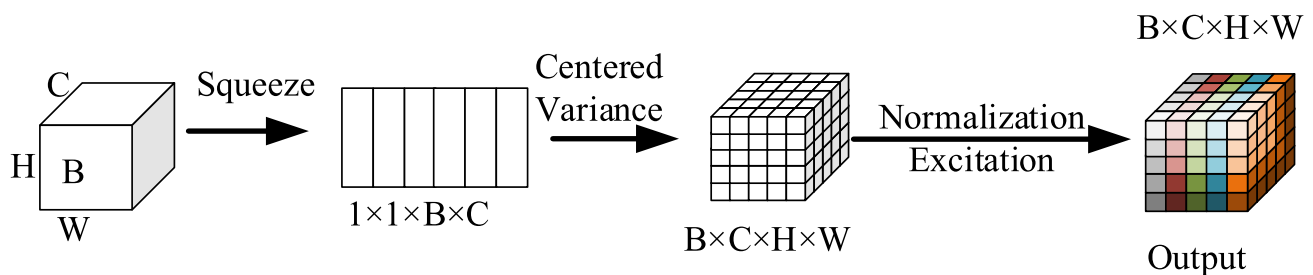


Fig.12 Similarity Attention Module

$$y_c^{(h,w)} = \frac{\tilde{x}_c^{(h,w)}}{4\left(\frac{1}{n_{h,w}}\sum \tilde{x}_c^{(h,w)} + \lambda\right)} + 0.5 \quad (12)$$

$$w_c^{(h,w)} = \sigma(y_c^{(h,w)}) = \frac{1}{1 + e^{-y_c^{(h,w)}}} \quad (13)$$

$$\hat{x}_c^{(h,w)} = w_c^{(h,w)} \odot x_c^{(h,w)} \quad (14)$$

In (10)–(14), $x \in \mathbb{R}^{B \times C \times H \times W}$ is the input feature tensor, $\mu_c \in \mathbb{R}^{B \times C \times 1 \times 1}$ represents the mean of channel c , $\bar{x} \in \mathbb{R}^{B \times C \times H \times W}$ is the centralized variance tensor, and $\lambda = 10^{-4}$ is a smoothing parameter to prevent division by zero. The operation \odot denotes element-wise multiplication, and the output \hat{x} provides enhanced feature representation.

Compared with existing attention-based object detection models such as YOLO-SPDConv and YOLO-ECANet, the proposed YOLO-ADHF-SimAM offers a targeted improvement for inland lake oil spill detection. YOLO-SPDConv enhances small-object detection through convolutional structural optimization, and YOLO-ECANet emphasizes efficient channel attention. In contrast, the ADHF module in this study models the physical diffusion characteristics of oil in water across multiple spatial scales, while the SimAM module provides spatially adaptive attention without introducing extra parameters. This combination addresses the spectral ambiguity and spatial fragmentation of oil spill targets in complex inland environments, resulting in higher precision and robustness under diverse conditions.

4 Analysis of Model Improvement Performance

4.1 Basic Training Parameters

The oil spill multispectral images were input into the improved model for training. The optimizer used in this experiment was SGD (Stochastic Gradient Descent), with a learning rate set to 0.01. The dataset underwent 200 training epochs. The input image dimensions for the model were 640×640 pixels, and in each update step, four images were processed simultaneously, as shown in Table 5.

4.2 Performance Evaluation Metrics

Precision (P): The proportion of correctly predicted positive samples among all predicted positives.

Recall (R): The proportion of correctly predicted positive samples among all actual positives.

Table 5 Network parameters and values

Parameter	Value
Optimizer	SGD
Learning rate	0.01
Epoch	200
Imgsz	640
Batch	4

AP (Average Precision): The area under the Precision-Recall curve.

mAP@50: The mean AP across all categories when the IoU threshold is 0.5.

mAP@50–95: The mean AP computed at IoU thresholds ranging from 0.5 to 0.95.

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \int_0^1 P(R) dR \quad (17)$$

$$mAP_{50} = \frac{1}{N} \sum_{i=1}^N AP_{i, IoU=0.5} \quad (18)$$

$$mAP_{50-95} = \frac{1}{10} \sum_{k=0}^9 mAP_{50+k \times 0.05} \quad (19)$$

In (15)–(19), within the context of object detection tasks, True Positive (TP) refers to correctly identified targets, i.e., objects that exist and are accurately detected; False Positive (FP) represents incorrectly detected targets, i.e., objects that do not exist but are misclassified as positive; and False Negative (FN) indicates existing targets that were not detected. N denotes the total number of object categories in the detection task. The mean Average Precision (mAP), which is the average precision across multiple categories, serves as a key metric for evaluating the overall performance of the detection model, reflecting its effectiveness in detecting objects across various categories.

4.3 Prediction Results and Analysis

The dataset was randomly split into training, validation, and test sets in a 5:2:2 ratio, containing 347, 138, and 138 images, respectively. The images were fed into different models for training and learning, and the results are shown in Figs. 13 and 14.

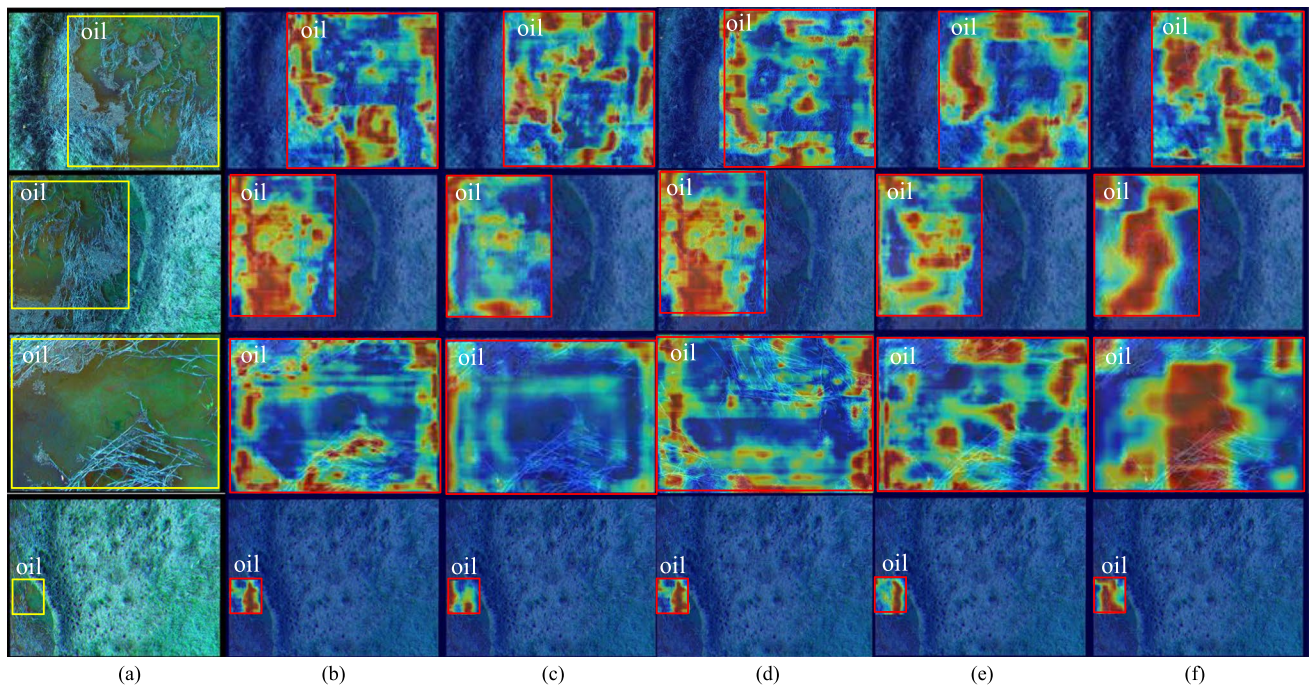


Fig.13 Results of ablation experiments, (a) is the original image; (b) is the YOLOv11 prediction result; (c) is the YOLO-HFAM prediction result; (d) is the YOLO-SimAM prediction result; (e) is the

YOLOADHF prediction result; (f) is the YOLO-ADHF- SimAM Network model(Ours) prediction result

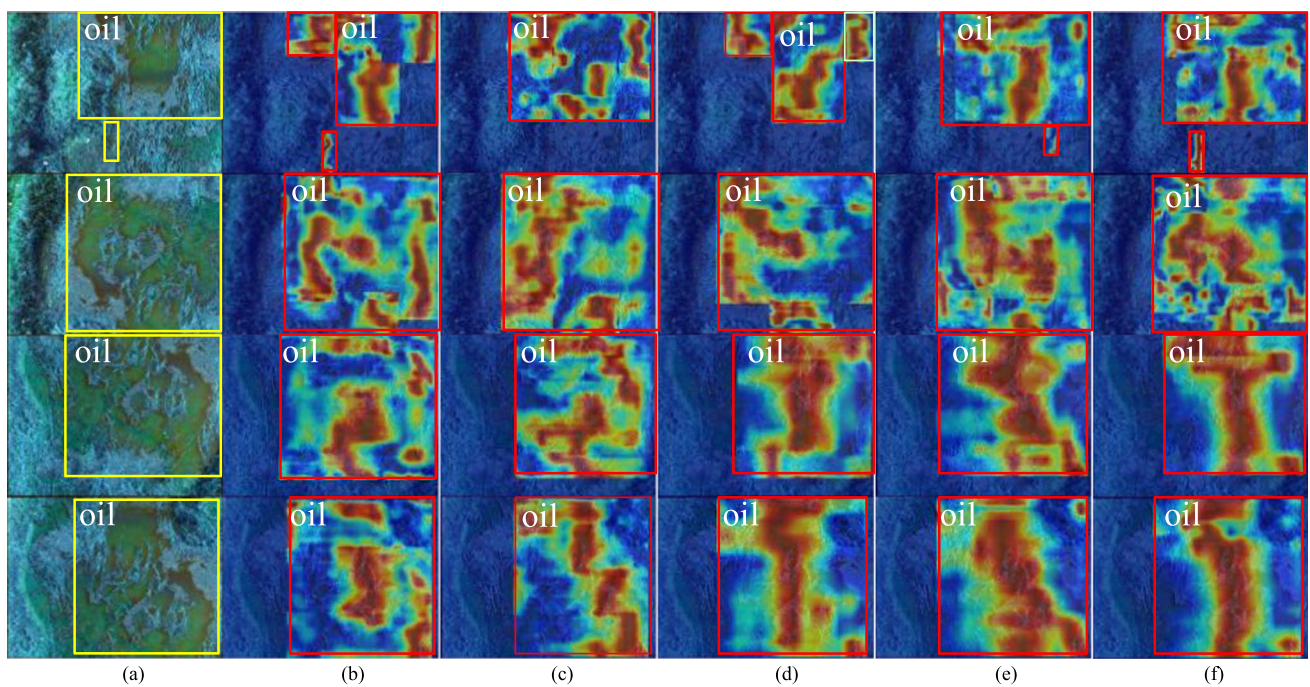


Fig.14 Comparative experiments, (a) is the original image; (b) is the YOLO-ECANet prediction result; (c) is the YOLO-CAFM prediction result; (d) is the YOLO-SPDConv prediction result; (e) is the YOLO-

KAN prediction result; (f) is the YOLO-ADHF-SimAM Network model(Ours) prediction result

Figures 13 and 14 illustrate the detection performance and attention distribution of different models for oil spill identification. The yellow bounding boxes indicate the ground truth, while the red boxes represent the predicted results. The heatmaps are generated based on the relative intensity of activation, using the mean value of each sample's gene expression level as a baseline. Regions with expression levels above the mean are considered positive and are marked in red, while those below the mean are negative and marked in blue. The color intensity reflects the magnitude of deviation from the mean, thus visually indicating how strongly a model responds to specific features within the oil-contaminated areas. This allows for intuitive evaluation of each model's focus and detection accuracy across varying target sizes and background complexities.

To comprehensively evaluate the performance of the proposed YOLO-ADHF-SimAM model, both comparative experiments with existing state-of-the-art methods and detailed ablation studies were conducted. The comparative experiments aim to benchmark the model against various advanced YOLO-based detection frameworks, while the ablation experiments are designed to isolate and assess the individual contributions of the SimAM attention mechanism, the HFAM module, and the proposed ADHF module. Key performance metrics such as Precision, Recall, mAP@50, and mAP@50–95 were employed to ensure a thorough and objective evaluation of detection accuracy and robustness in complex inland lake environments.

In addition to its detection accuracy, the YOLO-ADHF-SimAM model also inherits the inherent advantages of the YOLO architecture—high inference speed and deployment efficiency—making it particularly suitable for real-time environmental monitoring tasks, such as onboard

UAV-based oil spill detection. Unlike Transformer-based frameworks (e.g., RT-DETR), which often involve higher computational complexity and longer inference times due to multi-head self-attention mechanisms and deeper model structures, YOLO-based models are highly optimized for fast, single-pass predictions with minimal latency. This enables seamless integration into resource-constrained platforms and edge devices, where low power consumption and fast response times are critical. As a result, the proposed model not only ensures high detection accuracy in complex inland lake environments but also provides practical advantages in terms of real-time processing and field deployment, which are essential for large-scale, automated environmental surveillance systems.

Based on the comparative experimental results in Table 6, the proposed oil spill detection algorithm shows notable advantages for identifying oil–water mixtures in terrestrial lakes. In terms of detection accuracy, it achieves a Precision of 92.0%, surpassing all YOLO-based reference algorithms. Compared to YOLO-DynamicConv (90.3%), it improves by 1.7 percentage points, and it exceeds the current best-performing YOLO-CPDConv (91.3%) by 0.7 percentage points, effectively reducing false detections in complex aquatic environments. Although the Recall (82.8%) is slightly lower than YOLO-ECANet (83.9%) and YOLO-BRA (83.1%), the innovative precision-recall balancing mechanism narrows the Recall gap to within 1.1% while achieving a 3.1% and 3.9% increase in Precision over YOLO-ECANet (88.9%) and YOLO-BRA (88.1%), respectively. This is particularly valuable for environmental monitoring scenarios that demand high-confidence detections. The algorithm reaches 91.9% in mAP@50, marginally outperforming YOLO-CPDConv (91.8%) under the conventional IoU threshold of 0.5. Under

Table 6 Comparative experiment results of performance indicators

	P/%	R/%	mAP@50/%	mAP@50–95/%
YOLO- DynamicConv(Chen et al. 2020)	90.3	81.4	88.5	69.9
YOLO-ECANet(Wang et al. 2020)	88.9	<u>83.9</u>	90.9	72.2
YOLO-CGNet (Wu et al. 2021)	88.3	83.0	89.1	71.4
YOLO-DAttention(Xia et al. 2022)	89.5	79.4	90.9	72.7
YOLO-BRA (Zhu et al. 2023)	88.1	83.1	91.6	72.2
YOLO-CAFM(Zhou et al. 2023)	90.3	81.1	89.8	71.0
YOLO-LWA(Jamali et al. 2023)	88.3	78.5	89.1	68.4
YOLO-MLCA(Wan et al. 2023)	88.7	81.4	89.5	70.2
YOLO-SPDConv (Sunkara and Luo 2023)	88.4	82.2	91.8	<u>73.6</u>
YOLO-AFPN(Yang et al. 2024)	84.5	74.3	82.7	62.7
YOLO-KAN(Liu et al. 2024a, b)	91.3	77.2	89.8	70.2
RT-DETR(Zhao et al. 2024)	79.3	82.4	<u>92.5</u>	64.7
YOLO-RepNcsPeLAN (Wang et al. 2025)	88.7	81.4	89.9	68.8
Ours	92.0	82.8	91.9	72.4

The entries in bold represent the results obtained from my experiments, while the underlined entries indicate the best results among the compared methods

the stricter mAP@50:95 metric, it achieves 72.4%, slightly below YOLO-CPDConv (73.6%) but on par with YOLO-CAFFM (72.4%), and 2.5 percentage points higher than YOLO-DynamicConv (69.9%), indicating strong adaptability across different IoU thresholds.

Additionally, the RT-DETR model, representing the transformer-based detection paradigm, achieves Precision of 92.5%, Recall of 82.4%, mAP@50 of 92.5%, and mAP@50:95 of 64.7%. While its mAP@50 performance is slightly higher than that of the proposed method, its relatively lower mAP@50:95 suggests reduced localization accuracy across varying IoU thresholds. This further highlights the robustness and practical value of the proposed YOLO-ADHF-SimAM in scenarios requiring both high precision and consistency across scales and overlap conditions.

To evaluate the effectiveness of each component in the proposed model, an ablation study was conducted, and the results are presented in Table 7. The baseline YOLOv11 model achieved a Precision of 90.2%, Recall of 76.8%, mAP@50 of 88.6%, and mAP@50–95 of 70.4%. Incorporating the SimAM attention mechanism (YOLO-SimAM) led to a noticeable improvement in Precision (91.6%) and Recall (78.8%), although a slight decline in mAP@50–95 (68.7%) was observed, possibly due to reduced spatial generalization.

The integration of the HFAM module (YOLO-HFAM), which focuses on hierarchical multi-scale feature extraction, significantly enhanced Recall to 83.9% and mAP@50 to 91.6%. However, its Precision dropped to 88.1%, indicating a higher false-positive rate. To overcome this limitation, the improved ADHF module (YOLO-ADHF) introduced an adaptive diffusion-based cross-scale fusion strategy, resulting in better overall balance—achieving 91.8% Precision, 81.9% Recall, and 72.4% mAP@50–95.

YOLO-ADHF-SimAM combines both the ADHF module and the SimAM attention mechanism. It delivers the best comprehensive performance among all configurations,

achieving 92.0% Precision, 82.8% Recall, 91.9% mAP@50, and 72.4% mAP@50–95. These results confirm that the joint optimization of adaptive hierarchical fusion and attention mechanisms not only enhances detection accuracy but also improves robustness and generalization, making the proposed model highly suitable for real-world oil spill detection tasks in complex inland lake environments.

4.4 Discussion

While the SimAM module improves both Precision and Recall when applied alone, a slight drop in mAP@50–95 is observed compared to the baseline YOLOv11. This can be attributed to SimAM's parameter-free attention mechanism, which emphasizes prominent spatial features by suppressing redundant background signals. However, in complex inland lake environments, oil spills often appear as thin, irregular, or partially occluded regions with low contrast. As a result, SimAM may unintentionally suppress these critical fine-scale features, reducing performance under higher IoU thresholds (e.g., mAP@75 to mAP@95), which contributes to the lower overall mAP@50–95.

In contrast, the HFAM and ADHF modules focus on multi-scale spatial fusion, enhancing fine-grained feature retention and improving generalization to various oil spill shapes and thicknesses. However, HFAM alone results in reduced Precision due to limited adaptability to noisy backgrounds. ADHF addresses this by incorporating adaptive diffusion, enabling a better balance between background suppression and feature retention. When SimAM and ADHF are combined, their strengths complement each other—SimAM refines feature focus, while ADHF ensures scale-aware context aggregation—leading to optimal performance across all metrics.

5 Conclusion

This study presents an advanced UAV-based multispectral sensing framework for oil spill detection in terrestrial lakes, integrating deep learning techniques to enhance detection precision and robustness. By developing the Hierarchical Feature Aggregation Module (HFAM) for multi-scale feature extraction, the Adaptive DASI-HFAM Fusion (ADHF) for cross-scale feature interaction, and the SimAM attention mechanism for refined feature representation, the proposed model effectively addresses challenges posed by water flow disturbances, vegetation occlusion, and complex environmental backgrounds.

Experimental results demonstrate that the proposed YOLO-ADHF-SimAM network achieves a Precision of 92.0%, a Recall of 82.8%, and a mAP@50 of 91.9%, outperforming state-of-the-art methods in detecting oil spills under real-world conditions. Compared to traditional detection approaches, this method

Table 7 Ablation experiment results of performance indicators

	P/%	R/%	mAP@50/%	mAP@50–95/%
YOLOv11 (Khanam and Hussain 2024)	90.2	76.8	88.6	70.4
YOLO-SimAM (Yang et al. 2021)	91.6	78.8	88.8	68.7
YOLO-ADHF	91.8	81.9	91.2	<u>72.4</u>
YOLO-HFAM	88.1	<u>83.9</u>	91.6	<u>71.2</u>
Ours	92.0	82.8	91.9	72.4

The entries in bold represent the results obtained from my experiments, while the underlined entries indicate the best results among the compared methods

significantly reduces false positives while maintaining high recall, making it particularly suitable for time-sensitive environmental monitoring and emergency response applications.

The integration of UAV-mounted multispectral imaging with deep learning-based object detection provides a highly efficient and scalable solution for large-area oil spill surveillance, overcoming the limitations of satellite-based and manual inspection methods. The ability to capture spectral reflectance data across six wavelengths (450–840 nm) enhances oil spill discrimination, ensuring rapid and reliable detection even in complex inland water bodies.

Future research will focus on expanding dataset diversity to improve generalization across different oil spill scenarios and optimizing computational efficiency for real-time deployment on edge computing platforms. Additionally, further exploration of domain adaptation techniques and self-supervised learning could enhance model adaptability to unseen environments. This study contributes to the ongoing advancement of remote sensing technologies, offering a robust and intelligent framework for precise oil spill detection, with significant implications for environmental protection and sustainable resource management.

Funding National Natural Science Foundation of China, 32371864, Jian Xing

Data Availability The data that support the findings of this study were obtained from a third party and are subject to confidentiality agreements. Therefore, these data are not publicly available but may be available from the corresponding author upon reasonable request and with permission of the data provider.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barabanova EA, Vytovtov KA, Gladkikh TY, Migachev AN (2023) Environmental Monitoring of Water Surface Pollution in the Visible Range by Using UAVs. *J Commun Technol Electron* 68(SUPPL 3):S388–S392
- Chen Z, Li F, Diao Y, Zhao W, Fan P (2024) Knowledge-embedded multi-layer collaborative adaptive fusion network: Addressing challenges in foggy conditions and complex imaging. *Journal of King Saud University - Computer and Information Sciences* 36(10)
- Fang J, Chen X, Zhao J, Zeng K (2024) A scalable attention network for lightweight image super-resolution. *Journal of King Saud University - Computer and Information Sciences* 36(8)
- Feng F, Hu Y, Li W, Yang F (2024) Improved YOLOv8 algorithms for small object detection in aerial imagery. *Journal of King Saud University - Computer and Information Sciences* 36(6)
- Fingas M, Brown C (2014) Review of oil spill remote sensing. *Mar Pollut Bull* 83(1):9–23
- Hu H, Huang Y, Da X, Zhang H, Gao W, Ni L, Pan Y (2021) Optimization of Energy Utilization in Cognitive UAV Systems. *IEEE Sens J* 21(3):3933–3943
- Jamali A, Roy SK, Bhattacharya A, Ghamisi P (2023) Local Window Attention Transformer for Polarimetric SAR Image Classification. *IEEE Geosci Remote Sens Lett* 20:1–5
- Jingchun Z, Eg SuG, Shahrizal Sunar M (2024) Low-light image enhancement: A comprehensive review on methods, datasets and evaluation metrics. *Journal of King Saud University - Computer and Information Sciences* 36(10)
- Lennon M, Babichenko S, Thomas N, Mariette V, Mercier G, Lisin A (2006) Detection and mapping of oil slicks in the sea by combined use of hyperspectral imagery and laser induced fluorescence. *EARSeL eProceedings* 5(1):120–128
- Li Y, Zeng W, Yu Q, Zhou Q (2011) Adverse Effect of Marine Oil Spills on Human Health and Ecosystem: A Review. *Asian Journal of Ecotoxicology* 6(4):345–351
- Liu Z, Li Y, Shuang F, Huang Z, Wang R (2024) EMB-YOLO: Dataset, method and benchmark for electric meter box defect detection. *Journal of King Saud University - Computer and Information Sciences* 36(2)
- Ma X, Xu J, Wu P, Kong P (2022) Oil Spill Detection Based on Deep Convolutional Neural Networks Using Polarimetric Scattering Information From Sentinel-1 SAR Images. *IEEE Trans Geosci Remote Sens* 60:1–13
- Rahman S, Rahman MM, Abdullah-Al-Wadud M, Al-Quaderi GD (2016) Shoyaib M (2016) An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing* 1:35
- Rutoh EK, Zhi Guang Q, Bahadar N, Raza R, Hanif MS (2024) GAIR-U-Net: 3D guided attention inception residual u-net for brain tumor segmentation using multimodal MRI images. *Journal of King Saud University - Computer and Information Sciences* 36(6)
- Shao J, Li G, Zhang Z, Ye J, Hu Z, Zhang Z, Wang S, Li J (2021) Measurements of diffusion characteristics of hydroxyl radical with laser-induced fluorescence at high temperature. *Optics Communications* 488
- Sun G, Huang W, Chen P, Gao S, Wang X (2018) Advances in UAV-based Multispectral Remote Sensing Applications. *Transactions of the Chinese Society for Agricultural Machinery* 49(3):1–17
- Sun L, Zhang Y, Ouyang C, Yin S, Ren X, Fu S (2023) A portable UAV-based laser-induced fluorescence lidar system for oil pollution and aquatic environment monitoring. *Optics Communications* 527
- Tadmor E, Nevet A, Yahav G, Fish A, Cohen D (2015) Dynamic multi-spectral imaging using the vertical overflow drain structure. *IEEE Sens J* 15(7):3967–3972
- Tao J, Wei J, Zhou H, Meng F, Li Y, Wang C, Zhou Z (2024) Enhanced prediction model of short-term sea surface wind speed: A multi-scale feature extraction and selection approach coupled with deep learning technique. *Journal of King Saud University - Computer and Information Sciences* 36(8)
- Wan D, Lu R, Shen S, Xu T, Lang X, Ren Z (2023) Mixed local channel attention for object detection. *Eng Appl Artif Intell* 123
- Wang Y, Zhang R, Wang Z, Xue P, Cui M, Li M (2023) Snapshot-Type Polarization Spectral Imaging Resolution Enhancement Based

- on Dual-Channel Complementary Microarray Coding Imaging Fusion. *IEEE Sens J* 23(15):17127–17132
- Wang B, Yang J, Liu S, Sun C, Ma Y, Zhang J (2024) Identification of marine oil spill types based on multi-source remote sensing data using multi-scale dynamic convolution model. *Int J Remote Sens* 46:1–25
- Wu T, Tang S, Zhang R, Cao J, Zhang Y (2021) CGNet: A Light-Weight Context Guided Network for Semantic Segmentation. *IEEE Trans Image Process* 30:1169–1179
- Xing J, Zhang Y, Li B, Cao ZS, Zhan CL, Liu Y, Hao MZ, Jiang ZH (2024) Terrestrial Oil Spills Recognition Based on Small Samples From UAV Multispectral Images. *IEEE Sens J* 24(22):37786–37799
- Xu J, Cui C, Feng H, You D, Wang H, Li B (2019) Marine radar oil-spill monitoring through local adaptive thresholding. *Environmental Forensics* 20(2):196–209
- Xu Y, Du W, Deng L, Zhang Y, Wen W (2024) Ship target detection in SAR images based on SimAM attention YOLOv8. *IET Commun* 18(19):1428–1436
- Yang J, Ma Y, Hu Y, Jiang Z, Zhang J, Wan J, Li Z (2022) Decision Fusion of Deep Learning and Shallow Learning for Marine Oil Spill Detection. *Remote Sensing* 14:666
- Yang J, Hu Y, Ma Y, Wang M, Zhang N, Li Z, Zhang J (2023) Combined Retrieval of Oil Film Thickness Using Hyperspectral and Thermal Infrared Remote Sensing. *Remote Sensing* 15:5415
- Yang J, Wang J, Hu Y, Ma Y, Li Z, Zhang J (2023) Hyperspectral Marine Oil Spill Monitoring Using a Dual-Branch Spatial-Spectral Fusion Model. *Remote Sensing* 15:4170
- Yang G, Lei J, Tian H, Feng Z, Liang R (2024) Asymptotic Feature Pyramid Network for Labeling Pixels and Regions. *IEEE Trans Circuits Syst Video Technol* 34(9):7820–7829
- Yin SL, Sun FH, Liu WJ, Bi ZJ, Liu QC, Tian ZS (2023) Remote Identification of Oil Films on Water via Laser-Induced Fluorescence LiDAR. *IEEE Sens J* 23(12):13671–13679
- Zhang N, Yang J, Liu S, Zhang J, Ma Y (2024) Hyperspectral remote sensing identification of marine oil spills and emulsions using feature bands and double-branch dual-attention mechanism network. *Journal of Oceanology and Limnology* 42:1–17
- Zhou H, Luo F, Zhuang H, Weng Z, Gong X, Lin Z (2023) Attention Multihop Graph and Multiscale Convolutional Fusion Network for Hyperspectral Image Classification. *IEEE Trans Geosci Remote Sens* 61:1–14
- Zhu Q, Zhang Y, Li Z, Yan X, Guan Q, Zhong Y, Zhang L, Li D (2022) Oil Spill Contextual and Boundary-Supervised Detection Network Based on Marine SAR Images. *IEEE Trans Geosci Remote Sens* 60:1–10
- Cai X, Lai Q, Wang Y, Wang W, Sun Z, Yao Y (2024). Poly kernel inception network for remote sensing detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVF Conference on Computer Vision and Pattern Recognition*.
- Chen Y, Dai X, Liu M, Chen D, Yuan L, Liu Z (2020). Dynamic Convolution: Attention Over Convolution Kernels. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Khanam R, Hussain M (2024) YOLOv11: An Overview of the Key Architectural Enhancements. *ArXiv. abs/2410.17725*
- Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljai M, Hou T Y, Tegmark M (2024) KAN: Kolmogorov-Arnold Networks.
- Moskolai W R, Abdou W, Dipanda A, Kolyang (2021) Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review. *Remote Sensing*. 13 (23),
- Nahiyoon S A, Ren Z, Wei P, Li X, Li X, Xu J, Yan X, Yuan H (2024) Recent Development Trends in Plant Protection UAVs: A Journey from Conventional Practices to Cutting-Edge Technologies-A Comprehensive Review. *DRONES*. 8 (9),
- Nasayreh A, Jaradat A S, Gharaibeh H, Dawaghreh W, Mehamad Al Mamlook R, Alqudah Y, Al-Na'amneh Q, Sh. Daoud M, Migdady H, Abualigah L (2024) Jordanian banknote data recognition: A CNN-based approach with attention mechanism. *Journal of King Saud University - Computer and Information Sciences*. 36 (4), 102038.
- Pärt S, Kankaanpää H, Björkqvist J-V, Uiboupin R (2021) Oil Spill Detection Using Fluorometric Sensors: Laboratory Validation and Implementation to a FerryBox and a Moored SmartBuoy. *Frontiers in Marine Science*. 8
- Sunkara R, Luo T (2023). No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. *Machine Learning and Knowledge Discovery in Databases*, Cham, Springer Nature Switzerland.
- Wang C-Y, Yeh I H, Mark Liao H-Y (2025). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *Computer Vision – ECCV 2024*, Cham, Springer Nature Switzerland.
- Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,
- Xia Z, Pan X, Song S, Li L E, Huang G (2022). Vision Transformer with Deformable Attention. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,
- Xu S, Zheng S, Xu W, Xu R, Wang C, Zhang J, Teng X, Li A, Guo L (2024). HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection. *2024 IEEE International Conference on Multimedia and Expo (ICME)*,
- Yang L, Zhang R-Y, Li L, Xie X (2021). SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *International Conference on Machine Learning*,
- Zhao Y, Lv W, Xu S, Wei J, Wang G, Dang Q, Liu Y, Chen J (2024). DETRs Beat YOLOs on Real-time Object Detection. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,
- Zhu L, Wang X, Ke Z, Zhang W, Lau R (2023). BiFormer: Vision Transformer with Bi-Level Routing Attention. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.