

Article

LCSC-UAVNet: A High-Precision and Lightweight Model for Small-Object Identification and Detection in Maritime UAV Perspective

Yanjuan Wang ¹, Jiayue Liu ¹, Jun Zhao ¹, Zhibin Li ¹, Yuxian Yan ², Xiaohong Yan ¹, Fengqiang Xu ¹
and Fengqi Li ^{1,*}

¹ College of Rail Intelligent Engineering, Dalian Jiaotong University, Dalian 116028, China; wangyanjuan@djtu.edu.cn (Y.W.); liujiayue117@163.com (J.L.); 17741227783@163.com (J.Z.); lzb314314@163.com (Z.L.); yanxiaohong@djtu.edu.cn (X.Y.); xfq@djtu.edu.cn (F.X.)

² Information Science and Technology College, Dalian Maritime University, Dalian 116026, China; yyx13287345718@163.com

* Correspondence: fengqi-li@outlook.com

Abstract: Unmanned Aerial Vehicle (UAV) object detection is crucial in various fields, such as maritime rescue and disaster investigation. However, due to small objects and the limitations of UAVs' hardware and computing power, detection accuracy and computational overhead are the bottleneck issues of UAV object detection. To address these issues, a novel convolutional neural network (CNN) model, LCSC-UAVNet, is proposed, which substantially enhances the detection accuracy and saves computing resources. To address the issues of low parameter utilization and insufficient detail capture, we designed the Lightweight Shared Difference Convolution Detection Head (LSDCH). It combines shared convolution layers with various differential convolution to enhance the detail capture ability for small objects. Secondly, a lightweight CScConv module was designed and integrated to enhance detection speed while reducing the number of parameters and computational cost. Additionally, a lightweight Contextual Global Module (CGM) was designed to extract global contextual information from the sea surface and features of small objects in maritime environments, thus reducing the false negative rate for small objects. Lastly, we employed the WIoUv2 loss function to address the sample imbalance issue of the datasets, enhancing the detection capability. To evaluate the performance of the proposed algorithm, experiments were performed across three commonly used datasets: SeaDroneSee, AFO, and MOBdrone. Compared with the state-of-the-art algorithms, the proposed model showcases improvements in mAP, recall, efficiency, where the mAP increased by over 10%. Furthermore, it utilizes only 5.6 M parameters and 16.3 G floating-point operations, outperforming state-of-the-art models such as YOLOv10 and RT-DETR.

Keywords: UAV; CNN; object detection; small objects; lightweight



Academic Editor: Pablo Rodríguez-González

Received: 27 December 2024

Revised: 23 January 2025

Accepted: 27 January 2025

Published: 29 January 2025

Citation: Wang, Y.; Liu, J.; Zhao, J.; Li, Z.; Yan, Y.; Yan, X.; Xu, F.; Li, F.

LCSC-UAVNet: A High-Precision and Lightweight Model for Small-Object Identification and Detection in

Maritime UAV Perspective. *Drones*

2025, 9, 100. [https://doi.org/](https://doi.org/10.3390/drones9020100)

10.3390/drones9020100

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

Unmanned Aerial Vehicles (UAVs) play a crucial role in various fields, especially in maritime search and rescue [1,2] and post-disaster assessments [3,4], where they offer significant advantages [5]. According to the European Maritime Safety Agency (EMSA) 2023 annual report, 2516 maritime accidents involving 2701 vessels were reported in 2022, highlighting the critical need for effective UAV-based detection systems. However, small-object detection in maritime scenes remains a significant challenge due to the altitude and angle of capture, which often result in small objects with degraded visual features,

directly impacting detection accuracy [6]. Dynamic sea surface changes, such as waves and light reflections, further complicate detection by causing object shape distortions, reduced edge contrast, and blurred contours, thereby decreasing detection precision and increasing the risks of missed and false detections [7]. Additionally, the size and weight limitations of UAVs restrict the use of more powerful computational resources [8], necessitating the development of lightweight, high-precision algorithms that can operate efficiently under hardware constraints.

With the development of deep learning technologies, researchers have turned to more advanced algorithms to address small-object detection from the UAV perspective [9]. Feature Pyramid Networks (FPNs) [10] improve small-object detection by fusing multi-scale features, but their high computational and memory requirements are not suitable for resource-constrained UAV platforms. Single-stage detection algorithms, such as the SSD [11], RetinaNet [12], and YOLO series (including YOLOv5 [13] and YOLOv7 [14]), are widely used due to their fast detection and simplified design. However, these algorithms still exhibit limitations in small-object localization, often leading to missed detections or location offsets. Attention mechanisms [15], such as EMA [16], CA [17], and GAM, enhance the network's focus on important regions but also add computational overhead. Federated learning (FL) [18,19] offers a distributed training solution for small-object detection, ensuring data privacy. However, it still faces challenges related to participant unreliability and computational efficiency, particularly on resource-constrained UAV platforms. Therefore, despite significant progress in deep learning methods, achieving lightweight models while maintaining high accuracy remains a challenge.

To address the challenges of UAV-based maritime object detection, this paper introduces LCSC-UAVNet, a lightweight model designed for such tasks. LCSC-UAVNet enhances detection accuracy while reducing computational complexity and memory usage, making it suitable for resource-limited UAVs. The model includes innovative modules: LSDCH for detail capture, CScConv for efficient feature extraction, CGM for global context to reduce missed detections, and WIoUv2 [20] loss for improved convergence and class imbalance handling. Experiments on the SeaDroneSee [21], AFO [22], and MOBdrone [23] datasets show that LCSC-UAVNet outperforms state-of-the-art algorithms, achieving mAPs of 82.3%, 95.8%, and 98.5%, respectively. With only 5.6 M parameters and 16.3 G FLOPs, LCSC-UAVNet is significantly more lightweight than DETR [24] and YOLOX [25], reducing parameters and computational load by 31.1 M, 86.64 G and 3.4 M, 10.5 G, respectively. These results highlight LCSC-UAVNet's high accuracy and efficiency for maritime rescue and UAV small-object detection.

The main contributions of this paper are as follows:

1. A novel, high-precision, and lightweight detection model designed for drone-based small-object detection over the sea was proposed;
2. To improve parameter utilization and enhance detail capture for small objects, we designed the LSDCH to effectively balance accuracy and efficiency in UAV-based maritime detection tasks;
3. A lightweight CScConv module was designed to enhance the performance of maritime small-object detection while reducing the number of parameters and computational cost;
4. A lightweight CGM was proposed to effectively capture global contextual information from the maritime environment and extract features pertinent to small objects within drone images.

2. Relative Works

2.1. Small-Object Detection

To address challenges in small-object detection with deep learning algorithms on UAV-mounted vision systems, researchers have conducted meticulous optimizations and adjustments. Li et al. [26] introduced the more accurate small-object detection (MASOD) structure. This structure enhances recognition accuracy in complex scenarios. Yi et al. [27] utilized a dual-branch attention mechanism to enhance local modules in the feature extraction network for small-object detection, while also using vision transformer blocks to maximize the representation of feature maps to improve detection accuracy. However, the lightweight aspect still needs improvement. Chen et al. [28] designed a Super Feature Pyramid Network for marine remote sensing object detection. By highlighting the features of small objects and suppressing background features, they improved the ability to detect small objects. Nevertheless, this method increases model complexity and may impact both training and inference times.

2.2. Lightweight

Responding to the pressing need for lightweight models in the field of computer vision, researchers are focusing on developing more compact and efficient modules. Xie et al. [29] introduced the Partial Hybrid Dilated Convolution (PHDC) block to make models lightweight. This module combines hybrid dilated convolution and partial convolution to increase the receptive field at a lower computational cost. However, it has certain shortcomings in detecting small objects. Ge et al. [30] introduced a lightweight CNN–Transformer network that effectively extracts both local and global features. However, challenges remain in accurately detecting small objects. Lastly, Chen et al. [31] developed the LODNU network, a specialized YOLOv4-based solution for UAV vision tasks. Employing depthwise separable convolutions, this network efficiently reduces the parameter count. However, further refinements are needed to enhance the overall efficacy of the network, particularly in the precision and speed of small-object detection.

2.3. Maritime Object Detection in Complex Environments

Confronting the intricacies of object detection in drone imagery, exacerbated by multifaceted backgrounds and erratic lighting, researchers have pursued diverse strategies. Pikun et al. [32] unveiled a drone object detection framework. Leveraging advanced image dehazing methodologies, they combated the visual obstructions posed by maritime phenomena, notably sea fog. Wang et al. [33] enriched their model with a channel attention mechanism. This innovation alleviated the obscuration of objects amidst dense background clutter during the detection and tracking phases. Yet, the enhancement of small-object detection accuracy remains a frontier for further exploration. Zeng et al. [34] engineered an enhanced Selective Enhancement Block (SEB) module. Its primary objective was to sharpen the differentiation between foreground entities and background elements. Notable strides in compressing model dimensions were not achieved.

3. Method

3.1. Overall Network Architecture of the Model

The overall architecture of the proposed model, LCSC-UAVNet, is illustrated in Figure 1 and comprises three main components: Backbone, Neck, and Head. The Backbone consists of an alternating structure of five CBS convolutional modules, four CGMs, and a single SPPF module. The lightweight CGM plays a pivotal role in this section, effectively extracting both global contextual information from the maritime environment and small-object features within the images. Transitioning to the Neck, this section alternates between

two CScConv modules, three CGMs, and two CBS convolutional modules. The CScConv module integrates the robust capabilities of CNNs [35] with lightweight design principles, enhancing the accuracy of small-object detection in maritime settings while reducing both the number of parameters and computational costs. In the Head section, three LSDCH detection heads are employed. These heads utilize shared convolutional layers and various differential convolution techniques to improve parameter efficiency and enhance feature extraction sensitivity. LSDCH detection heads are particularly adept at detecting small objects across multiple scales in complex maritime environments, providing accurate and stable results at different scales. Finally, the model utilizes WIoUv2 and classification losses to optimize training and learning from UAV imagery. This strategy ensures that the model adapts to the unique characteristics of UAV images, thereby improving detection accuracy and generalization. Overall, this architectural design not only enhances model performance but also increases its practicality for deployment in resource-constrained environments.

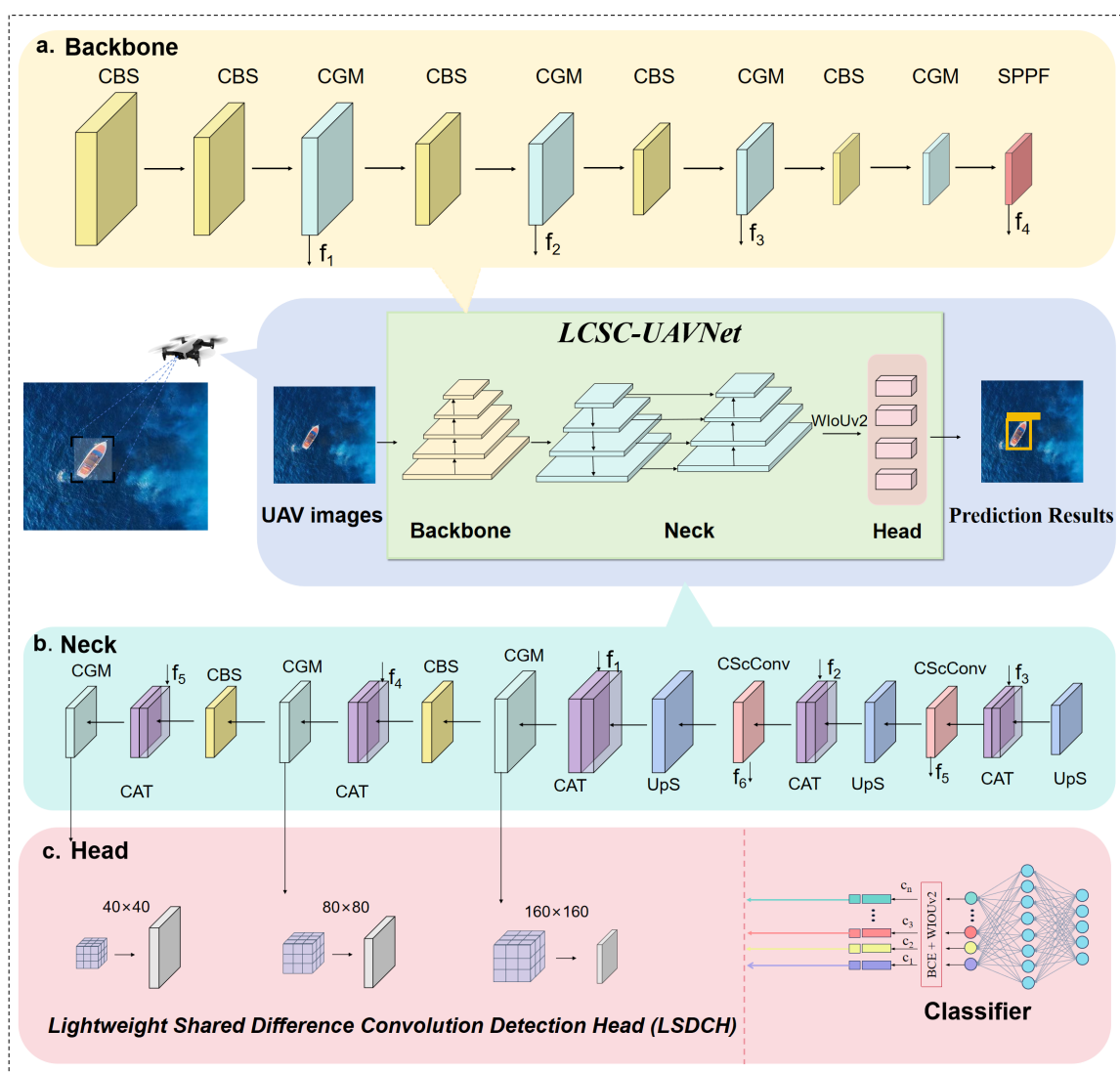


Figure 1. Overall network architecture of the LCSC-UAVNet model. (a) Backbone; (b) Neck; (c) Head.

3.2. Lightweight Shared Difference Convolution Detection Head (LSDCH)

In traditional detection head designs, independent detection heads are typically employed at different feature levels, resulting in low parameter utilization efficiency. To address this issue and enhance the model’s performance in handling complex maritime environments and multi-scale objects, we developed a new detection head called the

Lightweight Shared Difference Convolution Detection Head (LSDCH). The structure is illustrated in Figure 2 below. This detection head improves parameter utilization by sharing convolutional layers while incorporating various differential convolution techniques to enhance the capture of object details.

First, the input feature map is processed through an independent 1×1 convolutional layer to extract key features of the objects and adjust the number of channels. To ensure that features of varying sizes and complexities are represented effectively within the model, we employ Group Normalization in the convolutional layers. To further enhance the detail perception capability of small-object features, the LSDCH utilizes Shared Difference Convolution (SDConv) as its primary convolutional operation. SDConv incorporates four specialized convolution kernels: Central Difference Convolution (CDC), Angular Difference Convolution (ADC), Horizontal Difference Convolution (HDC), and Vertical Difference Convolution (VDC). These convolution kernels increase the weighting in the central, angular, horizontal, and vertical directions compared to standard convolutions, allowing for the capture of object detail features from multiple orientations and enhancing the model's ability to recognize object shapes and boundaries. The feature maps produced by SDConv undergo processing through Group Normalization and an activation function, resulting in more refined features. Consequently, the details within the feature maps are progressively enhanced through multiple convolutional operations, improving detection accuracy. Compared with other detection heads (such as the detection head of YOLOv5), LSDCH achieves model lightweighting by reducing the number of parameters and computational load through the use of shared convolutional layers instead of independent convolutional layers.

To optimize the issue of object scale in detection results, LSDCH integrates a learnable scale layer after the WIoUv2 loss function. After the boundary box predictions, each feature map is fine-tuned through a learnable scaling layer, further refining the predictions of the boundary boxes. This learnable scaling layer is instrumental in handling multi-scale objects, particularly in the context of UAV-based small-object detection in maritime scenarios, effectively addressing the challenges posed by complex maritime environments and variations in object scales within UAV imagery.

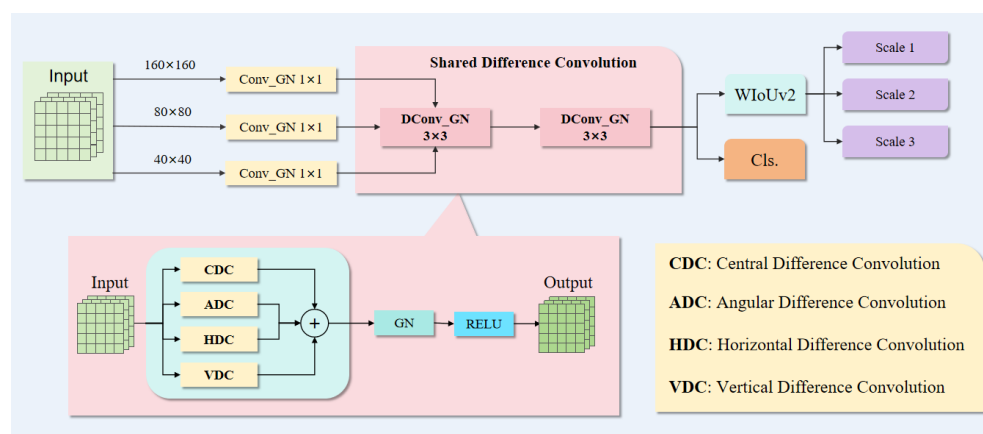


Figure 2. Overall network architecture of the LSDCH.

3.3. CScConv Module

CNNs are pivotal in deep learning, particularly for object detection. However, their high computational demands and large parameter counts pose challenges for resource-constrained systems like drones. To mitigate this, we proposed the CScConv module, as shown in Figure 3, a lightweight convolutional variant that improves efficiency while maintaining detection accuracy, essential for maritime surveillance and object detection

tasks. The CScConv module first applies a 1×1 convolution to the input feature X to adjust the number of channels, resulting in feature X_1 . The formula is expressed as follows:

$$X_1 = \text{Conv}_{1 \times 1}(X) \quad (1)$$

Feature map X_1 is divided into two parts by the spilt function in the CSP structure: one part, Y_1 , retains its original features, while the other part, Y_2 , is processed using the ScConv lightweight convolution unit. The mathematical representation of this segmentation operation is as follows:

$$Y_1, Y_2 = \text{Chunk}(X_1, \text{chunks} = 2, \text{dim} = 1) \quad (2)$$

ScConv [36] effectively compresses the architecture of convolutional neural networks (CNNs) and consists of two components: a Spatial Refinement Unit (SRU) and Channel Refinement Unit (CRU). First, the SRU module utilizes a gating mechanism to divide the input features into informative and non-informative parts, focusing computation on the informative portion while sparsely processing the non-informative part to reduce redundancy. The CRU module employs channel division to capture high-frequency details and low-frequency global features, combined with channel compression, grouped convolution, and pointwise convolution to reduce computational complexity while preserving feature interaction. This design enables the efficient processing of input features across both spatial and channel dimensions.

The concatenation of processed features Y_{SC} with the original features Y_1 is a critical operation within the CSP structure, enabling the effective fusion of features across different scales and levels, thereby yielding feature Y . The role of the CSP structure is to leverage features from different levels, thereby enhancing the model's feature extraction capabilities and robustness. Subsequently, a 1×1 convolution is applied to further integrate the channel information, yielding the final output features X_2 . The operational formula is expressed as follows:

$$Y = \text{Concat}(Y_1, Y_{SC}) \quad (3)$$

$$X_2 = \text{Conv}_{1 \times 1}(Y) \quad (4)$$

By effectively extracting and integrating spatial and channel features, CScConv not only improves the detectability of small objects but also enhances the algorithm's response speed and accuracy in real-time processing. This ultimately meets the efficiency demands of UAVs in maritime surveillance tasks. A quantitative evaluation of the CScConv module was conducted through ablation experiments.

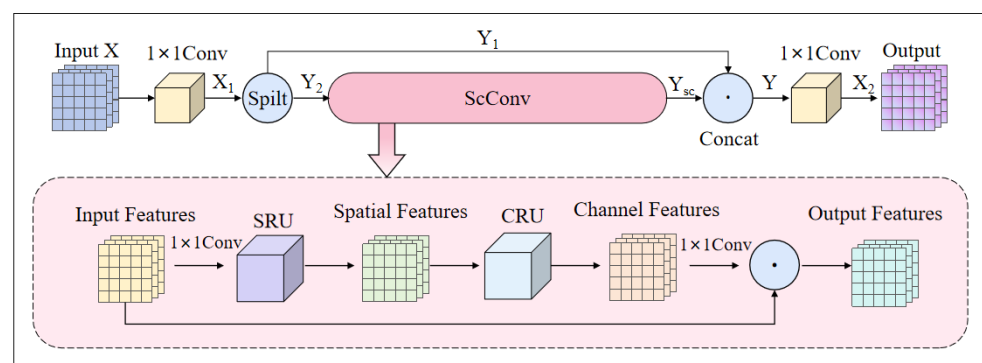


Figure 3. Overall network architecture of the CScConv.

3.4. Contextual Global Module (CGM)

In drone-captured aerial images, varying visual appearances, angles, scales, and lighting conditions complicate semantic understanding, particularly for visually similar but semantically distinct objects. To address these challenges, we introduce a lightweight Contextual Global Module (CGM) that leverages local, surrounding, and global contextual features. In enhancing global contextual information extraction, the CGM improves feature representation capabilities. Figure 4 presents a diagram of this module.

The operation of the CGM unfolds as follows: The input feature map X passes through a 1×1 convolutional layer, aimed at reducing the number of input data channels and thus mitigating subsequent computational complexity. The resultant feature map X_1 is then channeled into a Local Context Convolution Layer and a Surrounding Context Convolution Layer. The Local Context Convolution Layer comprises a 3×3 convolutional layer, utilizing padding to maintain the spatial dimensions of the feature map, thereby obtaining local features (X_{LF}). Compared to other convolutions, 3×3 convolutions offer a favorable balance between parameter count and receptive field, effectively capturing feature information while maintaining reasonable computational costs. As shown in the equation, where p represents the padding size, and d denotes the dilation rate,

$$X_{LF} = \text{Conv}_{3 \times 3}(X_1, p, N) \quad (5)$$

Conversely, the Surrounding Context Convolution Layer employs a dilated 3×3 convolutional layer with a dilation rate of 2, leveraging dilation rates to expand the receptive field, thereby capturing a broader spectrum of Surrounding Context Features (X_{SCF}). This configuration allows the convolutional layer to cover a larger area of the feature map without increasing the number of parameters. As shown in the equation, where N refers to the convolutional kernels that extract features from N channels,

$$X_{SCF} = \text{Conv}_{3 \times 3}(X_1, p, d, N) \quad (6)$$

The X_{LF} and X_{SCF} are concatenated along the channel dimension to form a Joint Feature (X_{JF}) representation that amalgamates local details with the environmental context. As shown in the equation,

$$X_{JF} = \text{Concat}(X_{SCF}, X_{LF}) \quad (7)$$

X_{JF} is subsequently fed into the Global Context Feature layer to generate Global Context Features (X_{GCF}). X_{JF} is compressed into a feature vector X_2 through an average pooling operation, followed by a fully connected layer that assigns weights to different channels, enhancing key features while suppressing irrelevant ones. The final output, denoted as $FC(X_2)$, is then added to X_{JF} , resulting in X_{GCF} , expressed as follows:

$$X_{GCF} = FC(X_2) + X_{JF} \quad (8)$$

UAV-based maritime small-object detection is a complex computer vision task, facing challenges such as background noise, varying illumination, occlusions, and diverse object sizes in practical applications. The CGM, integrated into the backbone and neck of the network, effectively distinguishes targets from background noise by fusing local and surrounding context features, thereby reducing false detections and missed detections. Additionally, it leverages global context features to refine the feature representation, enhancing the model's ability to accurately detect objects in complex maritime scenes. The quantitative evaluation of the CGM was conducted through ablation experiments.

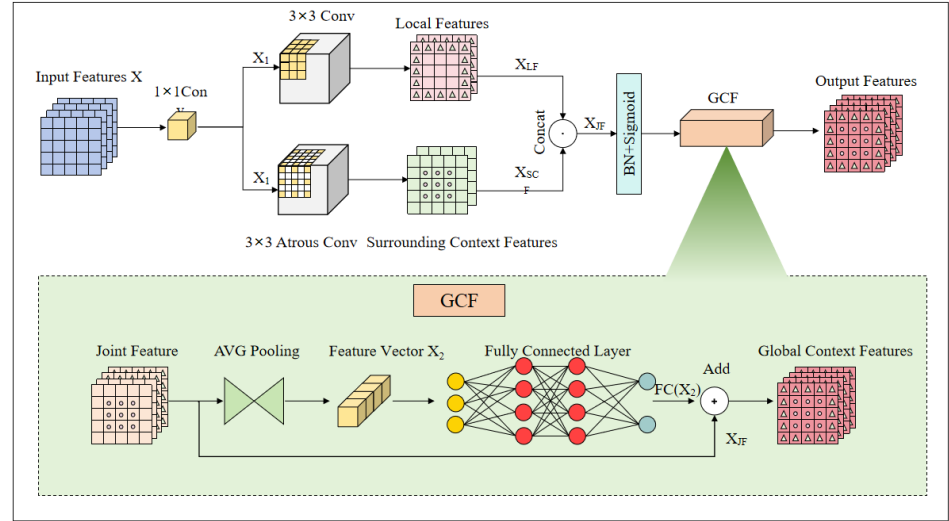


Figure 4. Overall network architecture of the CGM.

3.5. Wise-IoUv2

Due to the inconsistent altitude of drone-captured images, the object sizes vary, leading to significant errors and an imbalance of object samples. WIoUv2 is employed as the model's loss function to address this issue.

Three versions of the WIoU loss function (v1, v2, and v3) were tested in the experiment. After numerous trials, WIoU v2 was ultimately selected. For detailed experimental results, please refer to Supplementary Material Section S1. WIoU v1 was designed with an attention-based bounding box loss, while WIoU v2 and v3 incorporated a focusing mechanism by implementing a gradient gain calculation method.

WIoU constructs v1 based on the distance metric, which is represented by the following equation:

$$L_{WIoUv1} = L_{WIoU} L_{IoU} \quad (9)$$

$$L_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (10)$$

$$L_{IoU} = 1 - \frac{W_i H_i}{wh + w_{gt} h_{gt} - W_i H_i} \quad (11)$$

where x and y are the coordinates of the upper-left corner of the anchor frame, x_{gt} and y_{gt} are the coordinates of the upper-left corner of the object frame, W_g and H_g are the width and height of the minimum bounding box, W_i and H_i are the width and height of the area of overlap between the anchor frame and the object frame, W_{gt} and H_{gt} are the width and height of the object frame, and W and H are the width and height of the anchor frame, and $*$ is the separation of W_g and H_g from the computational graph. It is evident that WIoUv1 is a composite loss function consisting of two components: WIoU and IoU. WIoU represents the IoU value weighted by the distance between the predicted and ground truth boxes, while IoU refers to the ratio of the overlapping area to the non-overlapping area between the predicted and ground truth boxes.

WIoU v2 is a monotonic focusing mechanism for cross-entropy designed concerning Focal Loss, which effectively reduces the contribution of simple examples to the loss value. This allows the model to focus on difficult examples and gain improvement in classification performance. Building upon WIoUv1, gradient enhancement was introduced. The equation is as follows:

$$L_{WIoUv2} = L_{IoU}^{\gamma^*} L_{WIoUv1} \quad (12)$$

During the model training process, the gradient increment L_{IoU}^* became smaller and smaller as L_{IoU} decreased, leading to the problem of slow model convergence. To address this problem the mean value of $\overline{L_{IoU}}$ is proposed as a normalization factor,

$$L_{WIoUv2} = \left(\frac{L_{IoU}^*}{\overline{L_{IoU}}} \right)^\gamma L_{WIoUv1} \quad (13)$$

where $\overline{L_{IoU}}$ is the sliding average with momentum m , keeping the gradient gain $r = \left(\frac{L_{IoU}^*}{\overline{L_{IoU}}} \right)^\gamma$ overall high as a solution to the problem.

It introduces a weighting mechanism to better handle the overlapping areas of objects of different sizes, thereby providing more precise gradient signals. This is particularly important in maritime object detection, where the sizes of the objects can vary greatly, from small boats to large ships. Therefore, WIoUv2 is incorporated into the LCSC-UAVNet model to address the current issues of sample imbalance and varying object sizes.

4. Experimental Results and Discussion

To thoroughly assess the algorithm's generalizability and performance, a comprehensive set of experiments were conducted on three well-regarded maritime datasets: SeaDronesSee, AFO, and MOBdrone. Additionally, the algorithm was compared with current state-of-the-art detection methods, confirming its superior performance.

4.1. Dataset and Experimental Design

1. SeaDroneSee dataset: The SeaDronesSee dataset is a large-scale benchmark for visual object detection and tracking. Drones capture top-down images from 5 to 260 m high and at 0 to 90 degree angles, with details on height, angle, etc. This paper's dataset had 5630 images and 24,131 objects. Despite re-annotation due to sample imbalance, the original competition allocation was used: 2975 training, 859 validation, and 1796 testing images. To test algorithm robustness, Gaussian noise was randomly added to 45% of the validation images.

2. AFO dataset: The AFO dataset is a free dataset designed for training machine learning and deep learning models for maritime search and rescue applications. It includes aerial drone videos featuring 40,000 manually annotated maritime objects floating on the water, many of which are quite small in size, making them challenging to detect. The dataset was divided into three subsets in order to prevent the model from overfitting to the given dataset: the training set (64.7%), the validation set (13.48%), and the testing set (19.12%). To test algorithm robustness, Gaussian noise was randomly added to 45% of the validation images.

3. MOBDrone dataset: The MOBdrone dataset is a large collection of aerial images with 126,170 frames from 66 video clips taken by a drone at 10 to 60 m above sea level, featuring objects in five categories: people, boats, lifebuoys, surfboards, and wood. Due to the vast number of images, a random subset covering diverse scenes was selected for this study and split into a 7:2:1 ratio for training, validation, and testing maritime tasks. To test algorithm robustness, Gaussian noise was randomly added to 45% of the validation images. Specific training information for the three datasets is provided in Table 1.

The analysis results of small targets across three datasets are presented in Supplementary Material Section S2.

Table 1. Table of training information for the three datasets.

Datasets	Category	Epoch
SeaDroneSee	swimmer, floater, boat	300
AFO	human, windboard, boat, buoy, sailboat, kayak	300
MOBDrone	person, boat, surfboard, wood, lifebuoy	100

4.2. Experimental Environment Configuration

The PyTorch framework was used to implement the proposed method. To run the algorithm, a deep learning environment for PyTorch was set up on a Ubuntu 18.04 system, including a CPU with 15 vCPUs AMD EPYC 7642 48-Core Processor, 80 GB of RAM, and a graphics card RTX 3090 (24 GB) with 24 GB of video memory. Other experimental settings are shown in Table 2.

Table 2. Experimental setup parameters.

Parameters	Configuration
Learning rate	0.01
Momentum	0.937
Box	7.5
Works	8
Patience	50
Batch size	16
Image size	640 × 640

The evaluation metrics for the experiments in this study were the evaluation criteria for the object detection class. To assess the performance of the LCSC-UAVNet algorithm, the following metrics were utilized. Precision (P) represents the ratio of correctly detected samples to the total number of detected samples. This metric is calculated based on the number of true positives (TP) and false positives (FP) as follows:

$$P = \frac{TP}{TP + FP} \quad (14)$$

Recall (R) represents the ratio of correctly detected positive samples to the total number of samples in the test set. This metric is calculated based on the number of true positives (TP) and false negatives (FN) as follows:

$$R = \frac{TP}{TP + FN} \quad (15)$$

The precision–recall (P-R) curve is a plot with recall on the horizontal axis and precision on the vertical axis. Average precision (AP) is the average classification performance of a single category, defined as the area under the precision–recall (P-R) curve in the first quadrant. The equation is expressed as

$$AP = \int_0^1 P(R) dR \quad (16)$$

A higher AP value indicates that the model has better detection accuracy for that class. The value of mAP can be obtained from the area under the PR curve, which is the integral of the PR curve as follows:

$$mAP = \frac{1}{n} \sum_{i=1}^n (AP)_i \quad (17)$$

In the experimental process of deep learning object detection, it is often necessary to evaluate the model's suitability for practical applications. Therefore, metrics such as model parameters, detection speed, and other factors were also calculated.

Parameter count (Param) refers to the total count of model parameters during the training process. The more parameters a model has, the more complex it typically is. Therefore, a low parameter count is considered an important metric for evaluating model lightweighting.

Floating-point operations (FLOPs) refer to the mathematical computations a model performs during operation. This is a metric for measuring the computational load of a model, indicating how many floating-point operations can be performed per second.

Detection speed refers to the time required for a model to process an image and produce detection results. In practical applications, detection speed is crucial; the faster the detection speed, the higher the model's practicality.

4.3. Ablation Experiment

Ablation Experiments on the SeaDronesSee Dataset

To validate the effectiveness and improvements of the proposed modules, we used our self-developed network as the foundation, replacing its innovative modules with the C2f module to construct the baseline algorithm and conducted ablation experiments on the SeaDronesSee dataset. These experiments included training and testing phases to evaluate the contributions of different modules. The results of these ablation experiments are detailed in Table 3.

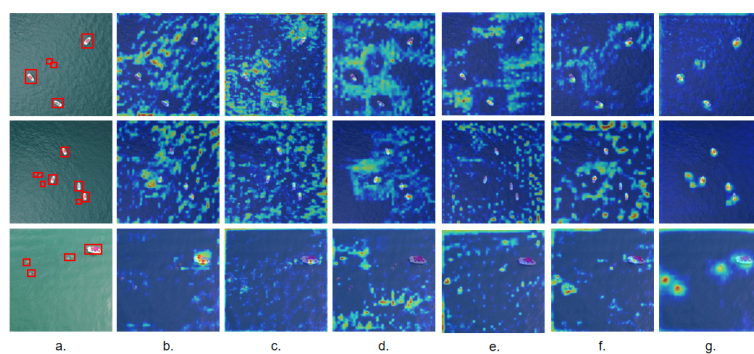
From the results of the ablation experiments, the following observations can be made: Ablation 1 provides the baseline algorithm results, serving as a reference for subsequent experiments. Ablations 2 to 5 demonstrate the results of individually introducing the four modules. Each module enhances the model's performance, with the addition of LSDCH in Ablation 2 showing improvements. The concept of shared convolutions reduces the floating-point operations by 10.8 G, achieving model lightweighting while also improving accuracy. In the comparison between Ablation 3 and Ablation 1, which involves ScConv and traditional convolution in the model, it can be observed that ScConv, under the influence of SRU and CRU, leads to an improvement in mAP@50. Additionally, the number of parameters is reduced by 0.2 G, and the floating-point operations are decreased by 1.1 G. The comparison between Ablation 4 and Ablation 1 highlights the performance differences between traditional convolution and the CGM. Quantitative analysis reveals that the CGM enhances the detection of small objects, with mAP@50 increasing by 2.3%, mAP@95 by 1.4%, precision by 0.7%, and recall by 1.9%. Additionally, the parameter count is reduced by 2.2 M, and the floating-point operations decrease by 7.2 G, further demonstrating that the CGM improves small-object detection accuracy while effectively reducing computational costs. Ablation 5 compared to Ablation 1 is the comparison between the WIoU and CIoU loss functions, achieving model performance enhancement on a balanced dataset without affecting computational complexity. Ablation 6 combines the LSDCH and CScConv modules. Compared to Ablation 2, there is a reduction in parameter count and floating-point operations, while all metrics show improvement. Ablation 7 integrates the LSDCH, CScConv, and CGM, further enhancing mAP@50 and improving accuracy and recall rates, indicating the effectiveness of the CGM in enhancing model performance. Ablation 8 includes all modules, showing an increase in precision and recall rates by 1.7% and 2.4%, respectively, compared to the baseline algorithm, with a reduction in parameter count by 2 M and in floating-point operations by 19.2 G. This highlights the balance between performance enhancement and computational efficiency.

Table 3. Table of results of ablation experiments on the SeaDroneSee dataset.

Ablation	LSDCH	CScConv	CGM	WIoU	mAP@50	mAP@95	P	R	Params	FLOPs
1	×	×	×	×	78.6%	45.6%	83.1%	72.4%	7.6 M	35.5 G
2	✓	×	×	×	80.4%	46%	84.6%	73.4%	8.0 M	24.7 G
3	×	✓	×	×	80.2%	45.6%	83.7%	72.8%	7.4 M	34.4 G
4	×	×	✓	×	80.9%	47%	83.8%	74.3%	5.4 M	28.3 G
5	×	×	×	✓	81.2%	46.2%	86%	73.4%	7.6 M	35.5 G
6	✓	✓	×	×	81.2%	46.5%	83%	74.3%	7.8 M	23.8 G
7	✓	✓	✓	×	81.6%	46.5%	84.6%	74.6%	5.6 M	16.3 G
8	✓	✓	✓	✓	82.3%	47.6%	84.8%	74.8%	5.6 M	16.3 G

These ablation experiments clearly show that the individual introduction of each module positively impacts model performance, with the LSDCH improving accuracy and recall. Moreover, the combination of modules brings about a performance boost. In particular, in Ablation 8, the model achieves optimal accuracy and recall while maintaining computational efficiency, which is highly valuable for real-world model deployment.

In the ablation experiments, feature analysis of the SeaDroneSee dataset's test set was conducted using heatmaps. Figure 5 illustrates the attention analysis results of the LCSC-UAVNet model, arranged from left to right as follows: (a) ground truth, (b) Ablation 1 heatmap, (c) Ablation 2 heatmap, (d) Ablation 3 heatmap, (e) Ablation 4 heatmap, (f) Ablation 5 heatmap, and (g) LCSC-UAVNet algorithm heatmap. The intensity of colors in the heatmaps indicates the model's focus on image regions during object detection predictions. Through comparison, it is evident that the LCSC-UAVNet model shows deeper colors in the heatmap where small objects are located, highlighting its ability to focus on small objects compared to the ablation algorithms. This visual analysis demonstrates that LCSC-UAVNet effectively identifies and prioritizes small-sized objects in images. The comparative curves illustrating the training and validation processes of the LCSC-UAVNet in conjunction with the baseline algorithm can be found in Supplementary Material Section S3.

**Figure 5.** Ablation experiment heatmaps on SeaDroneSee dataset. (a) ground truth, (b) Ablation 1 heatmap, (c) Ablation 2 heatmap, (d) Ablation 3 heatmap, (e) Ablation 4 heatmap, (f) Ablation 5 heatmap, (g) LCSC-UAVNet algorithm heatmap.

Visual comparison results of the ablation experiments on the SeaDroneSee dataset are shown in Figure 6. Two representative image scenarios were examined to evaluate the LCSC-UAVNet algorithm's performance in detecting small objects in complex maritime environments. As shown in Figure 6a, LCSC-UAVNet operates stably under medium-altitude and multi-category object recognition, accurately distinguishing various object types and showing improved accuracy and robustness compared to other ablation experiments that exhibited false positives. In Figure 6b, captured from a high altitude and facing challenges like complex sea surface textures and blurred edges of small objects, LCSC-UAVNet not

only accurately captures the ‘floaters’ category but also improves recognition accuracy compared to other ablation experiments.

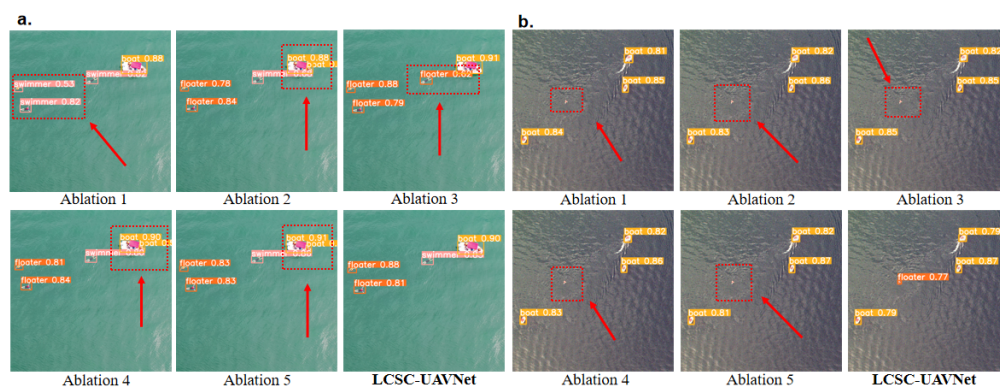


Figure 6. Visualization results of the ablation experiments on the SeaDroneSee dataset. (a) Medium-altitude UAV image; (b) high-altitude UAV image. The red arrow and dotted square in the figure indicate cases of missed or false detections by the comparison algorithm.

4.4. Comparison Experiment

4.4.1. Comparison Experiments on the SeaDroneSee Dataset

To assess the robustness of the proposed LCSC-UAVNet algorithm, it was evaluated on the SeaDroneSee dataset against state-of-the-art methods such as Faster R-CNN, DETR, YOLOv3, YOLOv3-tiny, YOLOv5, YOLOv6, YOLOX, YOLOv8, RT-DETR, and YOLOv10. The evaluation metrics included mAP@50, mAP@95, parameter count, computational requirements, recall, and detection speed. All methods were tested under consistent parameter settings and dataset scales to ensure a fair comparison. As shown in Table 4, LCSC-UAVNet outperformed these methods with lower computational and parameter counts while achieving higher precision and overall performance. Compared to efficient and robust YOLOv10, our algorithm improved mAP@50 by 15.6% and mAP@95 by 10%, reduced parameters by 1.6 M, decreased FLOPs by 5.3 G, increased recall by 15.4%, and sped up detection by 0.04 s.

Table 4. Table of results of experiments comparing methods on SeaDroneSee dataset.

Method	mAP@50	mAP@95	Params	FLOPs	Recall	Detection Speed
DETR	75.3%	38.6%	36.7 M	100.94 G	59.8%	1.09 s
Faster-RCNN	52.1%	33.8	40 M	206.7 G	30.3%	3.00 s
YOLOv3	77.2%	42.7%	61.52 M	77.54 G	70.3%	0.25 s
YOLOv3-tiny	67.7%	34.3%	8.9 M	12.9 G	65.1%	0.23 s
YOLOv5	73.1%	39.2%	20.85 M	47.9 G	69.8%	0.15 s
YOLOv6	65.2%	34.9%	11.5 M	27.6 G	62.1%	0.21 s
YOLOX	71.6%	42.3%	9.0 M	26.8 G	63.4%	0.24 s
YOLOv8	72%	42.8%	11.16 M	28.8 G	64.2%	0.17 s
YOLOv10	66.7%	37.6%	7.2 M	21.6 G	59.4%	0.12 s
RT-DETR	69.7%	36.7%	32.81 M	108 G	64.5%	0.15 s
LCSC-UAVNet	82.3%	47.6%	5.6 M	16.3 G	74.8%	0.08 s

In comparative experiments on the SeaDroneSee dataset, the LCSC-UAVNet model surpasses other state-of-the-art algorithms in key metrics such as precision, recall, mAP@50, and mAP@95. Figure 7a shows LCSC-UAVNet maintaining high precision (about 85%) in later training stages. In Figure 7b, despite similar smoothness in recall curves, LCSC-UAVNet leads by about 8 percentage points, indicating a high recall rate. Figure 7c demonstrates LCSC-UAVNet’s quick convergence and stable detection accuracy in mAP@50. Figure 7d further highlights LCSC-UAVNet’s advantages in mAP@95, confirming its excellence in small-object detection. Overall, LCSC-UAVNet shows outstanding performance in all key metrics, proving its advancement and practicality in this field.

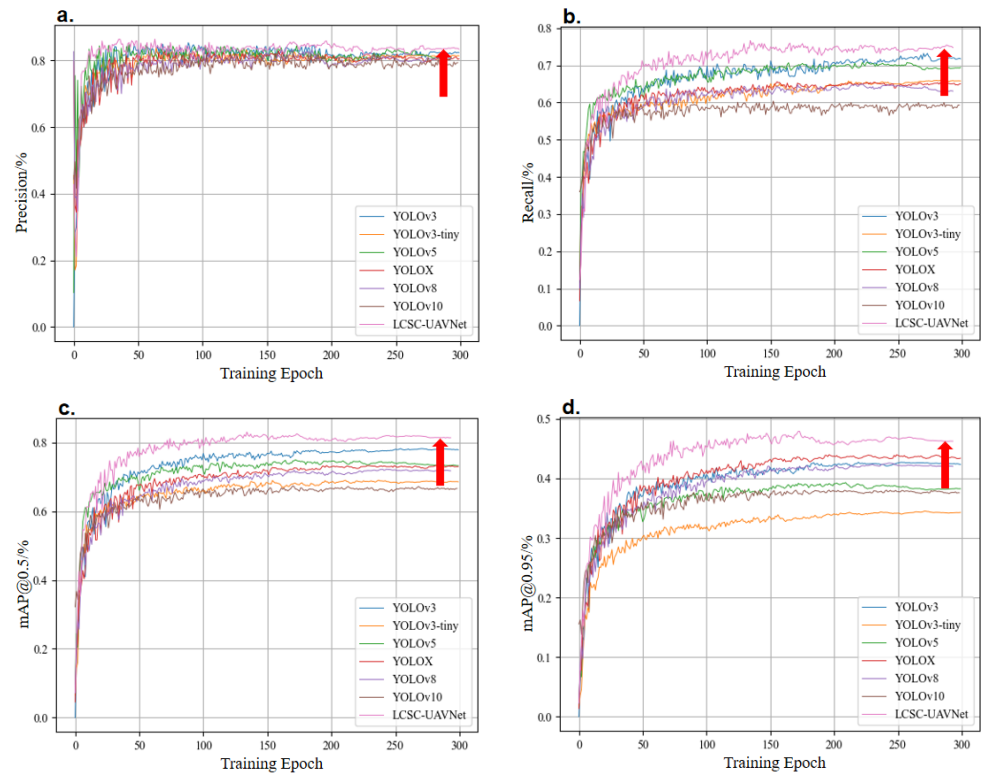


Figure 7. Comparison of the algorithm with the LCSC-UAVNet algorithm on the SeaDroneSee dataset for each metric. (a) Precision curve comparison, (b) recall curve comparison, (c) mAP@50 curve comparison, (d) mAP@95 curve comparison. The red arrow indicates the position of the proposed algorithm LCSC-UAVNet in the figure.

To validate the algorithm's performance under different maritime lighting conditions, this study selected three visualized result images from the SeaDroneSee dataset, as shown in Figure 8. Figure 8a represents a low-illumination UAV image, Figure 8b represents a UAV image under good lighting conditions, and Figure 8c depicts a dense small-target UAV image under low illumination and a simulated rainy environment, with the rain streaks generated using the model proposed in reference [37]. In the low-light environment of Figure 8a, where swimmer and buoy signals are weak and features are easily confused, other algorithms often suffer from missed detections or false positives. In contrast, LCSC-UAVNet accurately captures and distinguishes these objects, demonstrating its high recognition and differentiation capability under low-light conditions. In the high-altitude, well-lit scenario of Figure 8b, algorithms such as Faster R-CNN and YOLOv3 often miss detections of overlapping maritime human targets, whereas LCSC-UAVNet successfully identifies all expected objects, showing high detection completeness and accuracy. In the dense small-target scenario under low-light conditions with a simulated rainy environment, as shown in Figure 8c, occlusion, feature similarity, and raindrop interference often cause other algorithms to miss detections or misidentify raindrops as targets. In contrast, LCSC-UAVNet leverages local feature extraction and global context understanding to achieve precise detection without any missed targets. Overall, LCSC-UAVNet outperforms existing state-of-the-art algorithms in low-light, high-object-density, high-altitude, well-lit, and rainy maritime scenarios, achieving higher precision.

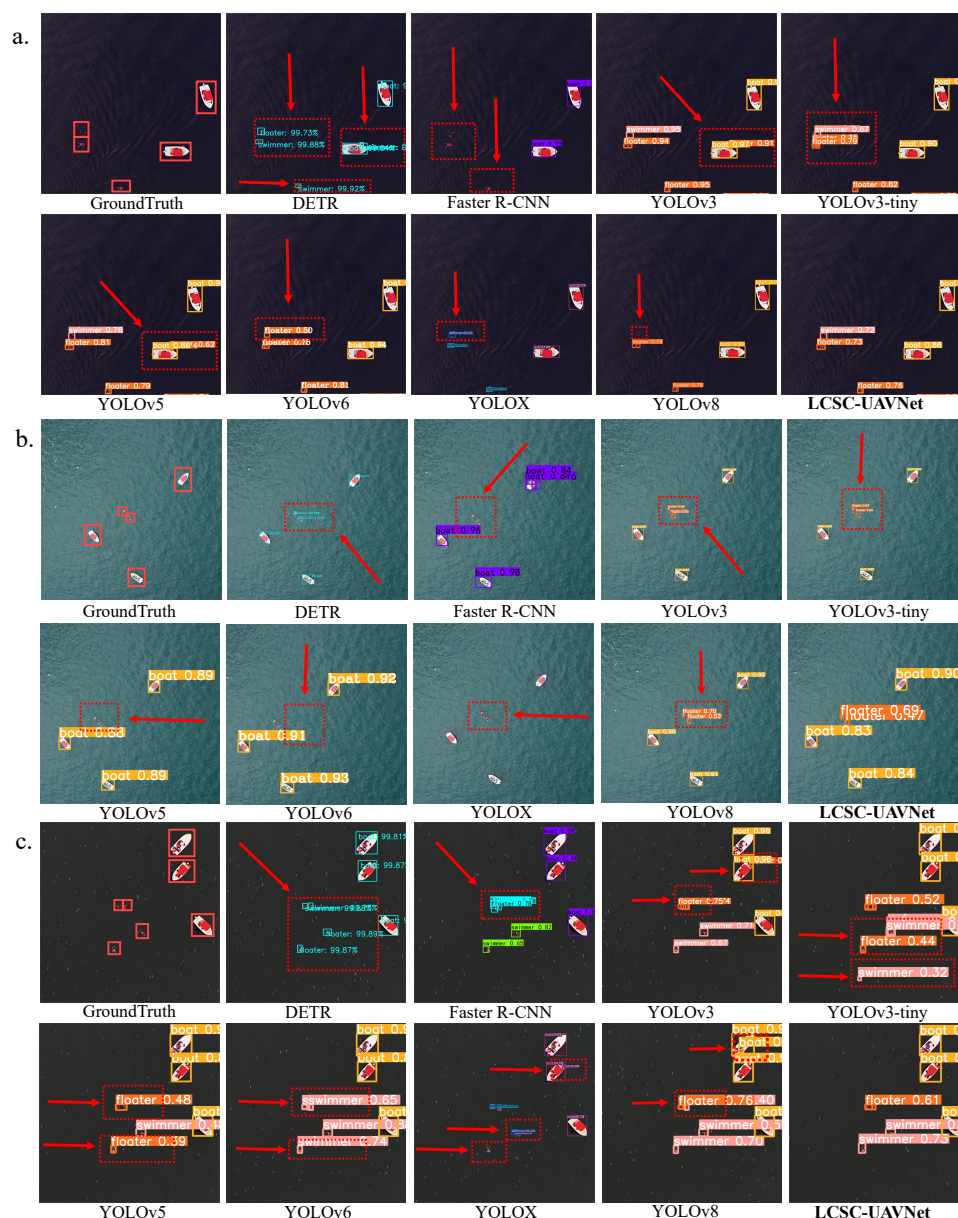


Figure 8. Visualization results of the comparative experiments on the SeaDronesSee dataset. (a) Low-illumination maritime environment, (b) high-altitude clear maritime environment, (c) low-illumination, dense small objects, and simulated rainy maritime environment. The red arrow and dotted square in the figure indicate cases of missed or false detections by the comparison algorithm.

4.4.2. Comparison Experiments on the AFO Dataset

In addition to on the SeaDronesSee dataset, experiments were conducted on the AFO public dataset to validate the proposed model's superiority further. Evaluation metrics were selected for assessment, including mAP@50, mAP@95, recall, and detection speed. The comparative experimental results on the AFO public dataset are presented in Table 5. The LCSC-UAVNet algorithm exhibits the lowest parameter count and computational complexity levels, further substantiating its efficiency and practicality. The LCSC-UAVNet model achieves a recall rate of 92.3%, outperforming both the YOLO series algorithms and two-stage algorithms like Faster R-CNN.

Table 5. Table of results of experiments comparing AFO datasets.

Method	mAP@50	mAP@95	Params	FLOPs	Recall	Detection Speed
DETR	93.3%	56.7%	36.7 M	100.94 G	73.4%	0.31 s
Faster-RCNN	64.7%	32.6%	40 M	206.7 G	53.7%	2.47 s
YOLOv3	94.7%	60.2%	61.52 M	77.54 G	90.2%	0.29 s
YOLOv3-tiny	90.7%	55.5%	8.9 M	12.9 G	87.6%	0.19 s
YOLOv5	94.2%	59.2%	20.85 M	47.9 G	91.7%	0.18 s
YOLOv6	86.7%	53.2%	11.5 M	27.6 G	71.3%	0.20 s
YOLOX	93.3%	63.5%	9.0 M	26.8 G	85.7%	0.19 s
YOLOv8	90%	60.1%	11.16 M	28.8 G	85.7%	0.19 s
YOLOv10	92%	63.1%	7.2 M	21.6 G	86.8%	0.17 s
RT-DETR	83.4%	44.9%	32.81 M	108 G	77.6%	0.25 s
LCSC-UAVNet	95.8%	64.4%	5.6 M	16.3 G	92.3%	0.16 s

When comparing the LCSC-UAVNet model with state-of-the-art algorithms in precision, recall, mAP@50, and mAP@95, LCSC-UAVNet shows performance improvements across all metrics. Figure 9 illustrates these comparisons, with Figure 9a for precision, Figure 9b for recall, Figure 9c for mAP@50, and Figure 9d for mAP@95. LCSC-UAVNet's precision curve is stable with higher accuracy and lower volatility. For recall, LCSC-UAVNet stabilizes earlier, after about 30 iterations, and is nearly 5 percentage points higher. In mAP@50, LCSC-UAVNet converges nearly twice as fast with a smoother curve. For mAP@95, LCSC-UAVNet maintains a higher value of 64.5%. Overall, LCSC-UAVNet excels in precise recognition, comprehensive recall, and average precision under varying accuracy requirements.

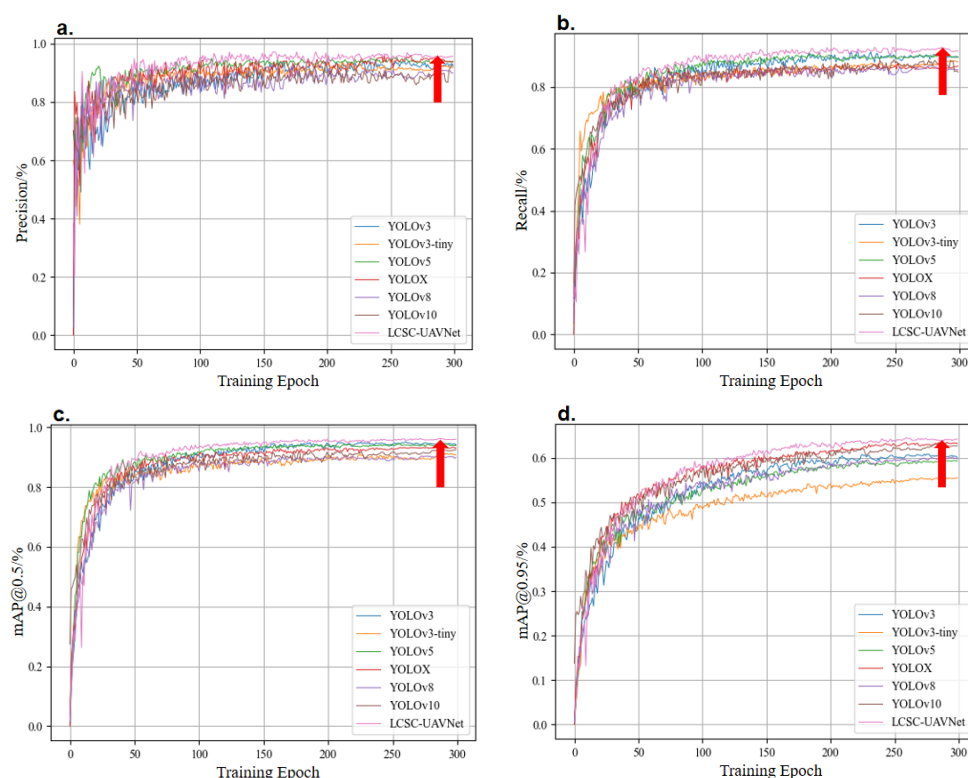


Figure 9. Comparison of the algorithm with the LCSC-UAVNet algorithm on the AFO dataset for each metric. (a) Precision curve comparison, (b) recall curve comparison, (c) mAP@50 curve comparison, (d) mAP@95 curve comparison. The red arrow indicates the position of the proposed algorithm LCSC-UAVNet in the figure.

To evaluate the algorithm's performance under different viewing angles and weather conditions, three representative visual results from the AFO dataset were selected, as shown in Figure 10. Figure 10a illustrates a UAV-captured scene in clear weather from a vertical viewpoint;

Figure 10b depicts a UAV-captured scene in overcast weather from an oblique viewpoint; and Figure 10c shows a UAV-captured scene with fog from a near-surface maritime perspective. In the vertical clear-weather perspective of Figure 10a, other algorithms exhibit false positives and missed detections, particularly for the buoy category. In contrast, LCSC-UAVNet accurately identifies the targets, demonstrating its capability for detecting small maritime objects from UAVs. In the oblique overcast perspective of Figure 10b, mainstream algorithms fail to detect the buoys accurately, whereas LCSC-UAVNet successfully identifies them, validating its high precision under oblique angles and low-light conditions. In Figure 10c, with the near-sea-level perspective and foggy conditions, distant swimmers are obscured by fog. LCSC-UAVNet is the only algorithm capable of accurately detecting these targets, proving its ability to capture small targets under such maritime conditions. LCSC-UAVNet incorporates design features such as differential convolutional detection heads and global context processing, which optimize the model's performance in complex maritime environments.

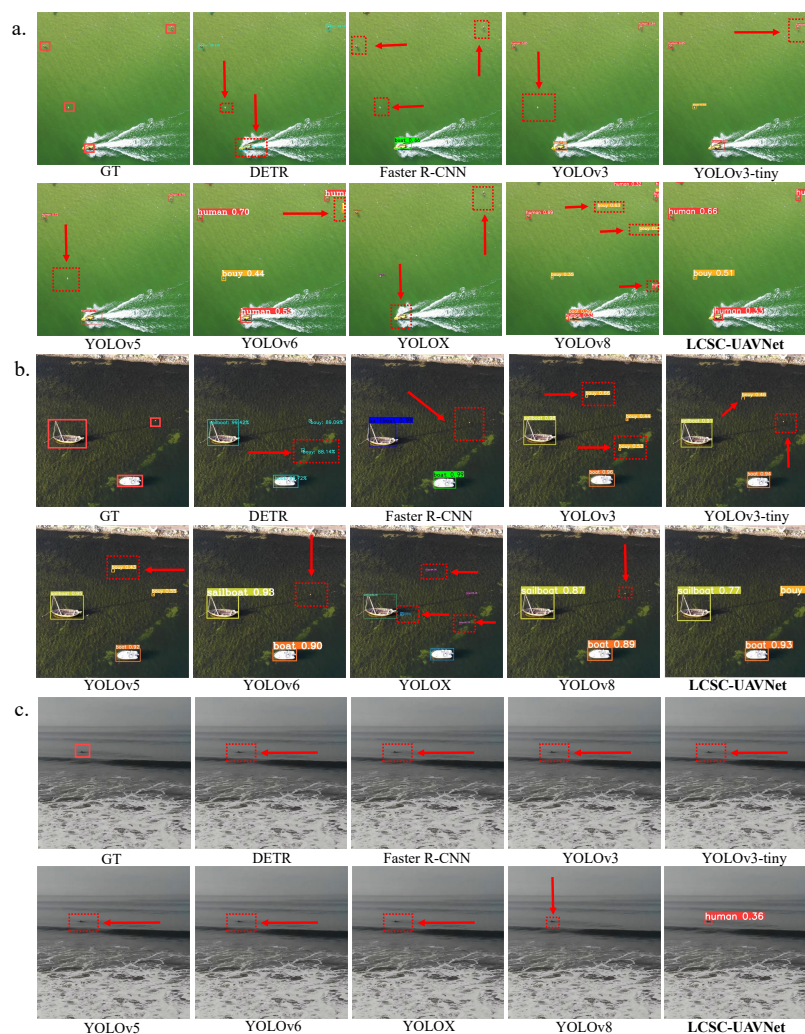


Figure 10. Visualization results of the comparative experiments on the AFO dataset. (a) is a UAV image captured from a vertical perspective under clear weather; (b) is a UAV image captured from an oblique perspective under overcast weather; (c) is a UAV image captured from a near-sea-level perspective under foggy conditions. The red arrow and dotted square in the figure indicate cases of missed or false detections by the comparison algorithm.

4.4.3. Comparison Experiments on the MOBDrone Dataset

The MOBdrone dataset was also utilized to further validate the generality and adaptability of the LCSC-UAVNet algorithm. MOBdrone, a large-scale publicly available mar-

itime object detection dataset, larger than datasets at the opposite end of the spectrum, served to further evaluate the feature extractability of this lightweight model. Comparative results with state-of-the-art object detection methods using the MOBDrone dataset are presented in Table 6. The LCSC-UAVNet algorithm demonstrates substantial improvements over several metrics compared to various algorithms.

Table 6. Table of results of experiments comparing MOBDrone datasets.

Method	mAP@50	mAP@95	Params	FLOPs	Recall	Detection Speed
DETR	88.6%	51.4%	36.7 M	100.94 G	72.3%	0.22 s
Faster-RCNN	65.9%	31.1%	40 M	206.7 G	70.7%	0.52 s
YOLOv3	97.3%	87.8%	61.52 M	77.54 G	96.3%	0.15 s
YOLOv3-tiny	92.9%	70.5%	8.9 M	12.9 G	91.7%	0.12 s
YOLOv5	94.2%	76.1%	20.85 M	47.9 G	93.6%	0.13 s
YOLOv6	88.2%	67.8%	11.5 M	27.6 G	81.3%	0.17 s
YOLOX	96.2%	84.7%	9.0 M	26.8 G	93.2%	0.16 s
YOLOv8	95.4%	77.0%	11.16 M	28.8 G	94.7%	0.12 s
YOLOv10	97%	81.7%	7.2 M	21.6 G	92%	0.10 s
RT-DETR	90.8%	70.6%	32.81 M	108 G	86.7%	0.22 s
LCSC-UAVNet	98.5%	87.5%	5.6 M	16.3 G	96.7%	0.08 s

To rigorously evaluate the algorithm's performance in complex sea states, this paper presents three visual results selected from the MOBDrone dataset, as shown in Figure 11. All images were captured at the same altitude to ensure a fair comparison. Figure 11a depicts a drone image of wave disturbances in complex sea states; Figure 11b shows a drone image of small targets on a calm sea surface; and Figure 11c shows a UAV image of small targets in a complex nearshore background under a simulated rainy environment, with the rain streaks generated using the model proposed in reference [37]. As shown in Figure 11a, on the image with wave disturbances, mainstream algorithms exhibit issues such as false negatives and bounding box overlaps in the detection of ships and personnel, while LCSC-UAVNet successfully avoids false detections. In Figure 11b, regarding the detection of two swimmers, mainstream algorithms encounter false positives or false negatives, struggling to handle variations in swimming posture, whereas LCSC-UAVNet accurately identifies the swimmers without any false detections or omissions. As shown in Figure 11c, in the task of detecting ships, onboard personnel, and lifebuoys under a simulated rainy environment, some algorithms detect only the ships while failing to recognize the personnel and lifebuoys. Other algorithms detect the personnel but fail to identify the lifebuoys and mistakenly classify raindrops as human targets. In contrast, LCSC-UAVNet successfully detects all targets, accurately identifying multiple small targets in complex scenarios. In conclusion, when confronted with complex sea surface textures, LCSC-UAVNet demonstrates robust target detection capabilities, accurately recognizing various targets within the images.

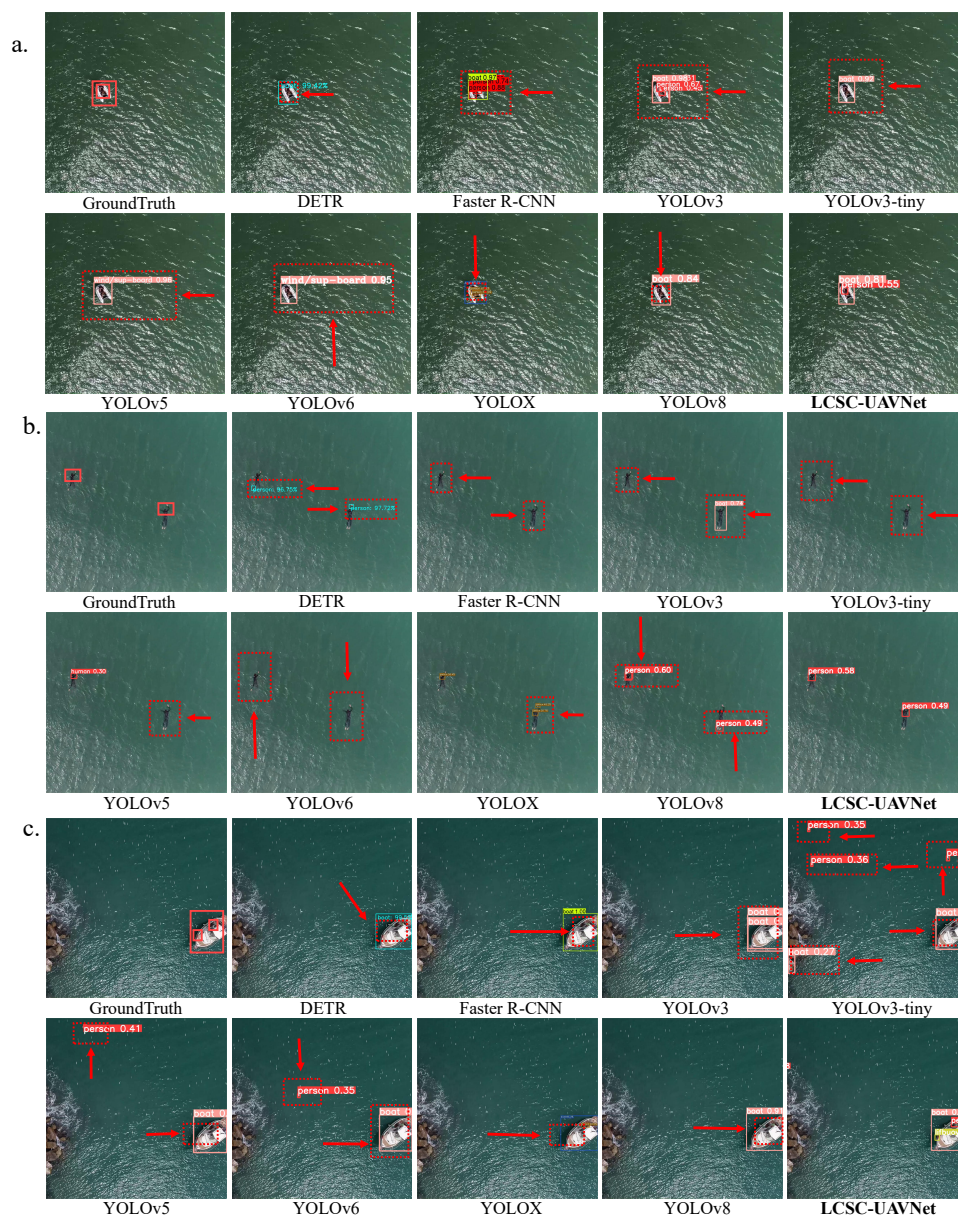


Figure 11. Visualization results of the comparative experiments on the MOBDrone dataset with complex sea states. (a) depicts a drone image of wave disturbances in complex sea states; (b) shows a drone image of small targets on a calm sea surface; and (c) illustrates a drone image of small targets in a complex nearshore background under a simulated rainy environment. The red arrow and dotted square in the figure indicate cases of missed or false detections by the comparison algorithm.

5. Conclusions

This study addresses the key challenges of low accuracy, slow processing speed, and inadequate lightweight design in maritime small-object detection. Through the development of the LCSC-UAVNet model, which integrates several innovative components including the Lightweight Shared Difference Convolution Detection Head (LSDCH), CSc-Conv, Contextual Global Module (CGM), and WIoUv2 loss function, the model enhances detection performance, especially for small objects in complex maritime environments. Experimental results demonstrate that LCSC-UAVNet outperforms existing state-of-the-art models on the SeaDroneSee, AFO, and MOBdrone datasets, validating its superior accuracy and computational efficiency.

Despite its achievements, the LCSC-UAVNet model still faces certain limitations. The model's ability to capture fine details of small objects requires further improvement,

constraining its detection accuracy and leading to suboptimal performance in complex environments. To address these challenges, future research will explore the integration of advanced techniques. Incorporating adaptive attention mechanisms (e.g., dynamic convolutions, multi-scale attention, and vision transformers) can enhance the model's robustness in complex backgrounds and dynamic scenarios. Leveraging generative adversarial networks (GANs) for data augmentation can simulate diverse scene characteristics, further enriching the training data. Techniques such as invariant learning (e.g., domain generalization) and adversarial domain adaptation (ADA) will be employed to improve the model's transferability across different scenarios, such as urban environments and complex weather conditions. Moreover, high-resolution feature fusion methods (e.g., BiFPN) and vision transformers (e.g., DETR) will be utilized to enhance the capture and representation of small-object features. Multi-task learning will also be explored to jointly optimize related tasks and strengthen feature sharing. The integration of these technologies will significantly enhance the LCSC-UAVNet model's detail-capturing ability and robustness, providing broader support for small-object detection in complex scenarios.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/drones9020100/s1>: Section S1: Evaluation of the Loss Function; Table S1: Table of experimental results for loss function comparison on SeaDroneSee dataset; Figure S1: Loss function training comparison curves on SeaDroneSee dataset; Section S2: Analysis of datasets; Table S2: Definition of large, medium, and small objects in the COCO dataset; Table S3: Distribution of large, medium and small objects in the three datasets; Section S3: Ablation experiment Supplement; Figure S2: Comparison curves of the training and validation process LCSC-UAVNet with the baseline algorithm.

Author Contributions: Conceptualization, Y.W.; methodology, J.L. and J.Z.; validation, Z.L.; formal analysis, X.Y. and Y.Y.; data curation, F.X.; writing—original draft preparation, J.L. and J.Z.; writing—review and editing, Y.W. and F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Natural Science Foundation of Liaoning Province (No.2024-MS-168), the Fundamental Research Funds for the Provincial Universities of Liaoning (No.LJ212410150030, LJ212410150024), the Research Foundation of Liaoning Province (No.LJKQZ20222447), and the Foundation of Yunnan Key Laboratory of Service Computing (Grant No.YNSC23118).

Data Availability Statement: The datasets used in the paper can be downloaded here “SeaDroneSee Dataset” at <https://seadronessee.cs.uni-tuebingen.de/> accessed on 11 March 2024, “AFO Dataset” at <https://www.kaggle.com/datasets/jangsienicajzkowy/afo-aerial-dataset-of-floating-objects?resource=download/> accessed on 11 January 2024, and “MOBDrone Dataset” at <https://aimh.isti.cnr.it/dataset/mobdrone/> accessed on 16 March 2024.

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

References

1. Alsamhi, S.H.; Shvetsov, A.V.; Kumar, S.; Shvetsova, S.V.; Alhartomi, M.A.; Hawbani, A.; Rajput, N.S.; Srivastava, S.; Saif, A.; Nyangaresi, V.O. UAV computing-assisted search and rescue mission framework for disaster and harsh environment mitigation. *Drones* **2022**, *6*, 154. [\[CrossRef\]](#)
2. Zhang, Y.; Tao, Q.; Yin, Y. A Lightweight Man-Overboard Detection and Tracking Model Using Aerial Images for Maritime Search and Rescue. *Remote Sens.* **2023**, *16*, 165. [\[CrossRef\]](#)
3. Sun, B.; Bi, K.; Wang, Q. YOLOv7-FIRE: A tiny-fire identification and detection method applied on UAV. *AIMS Math.* **2024**, *9*, 10775–10801. [\[CrossRef\]](#)
4. Xin, W.; Pu, C.; Liu, W.; Liu, K. Landslide surface horizontal displacement monitoring based on image recognition technology and computer vision. *Geomorphology* **2023**, *431*, 108691. [\[CrossRef\]](#)
5. Lyu, H.; Shao, Z.; Cheng, T.; Yin, Y.; Gao, X. Sea-surface object detection based on electro-optical sensors: A review. *IEEE Intell. Transp. Syst. Mag.* **2022**, *15*, 190–216. [\[CrossRef\]](#)

6. Yang, Z.; Yin, Y.; Jing, Q.; Shao, Z. A High-Precision Detection Model of Small Objects in Maritime UAV Perspective Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2023**, *11*, 1680. [\[CrossRef\]](#)
7. Zhang, Y.; Yin, Y.; Shao, Z. An Enhanced Target Detection Algorithm for Maritime Search and Rescue Based on Aerial Images. *Remote Sens.* **2023**, *15*, 4818. [\[CrossRef\]](#)
8. Zhang, L.; Zhang, N.; Shi, R.; Wang, G.; Xu, Y.; Chen, Z. SG-Det: Shuffle-GhostNet-based detector for real-time maritime object detection in UAV images. *Remote Sens.* **2023**, *15*, 3365. [\[CrossRef\]](#)
9. Wang, S.; Han, Y.; Chen, J.; Zhang, Z.; Wang, G.; Du, N. A deep-learning-based sea search and rescue algorithm by UAV remote sensing. In Proceedings of the 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC), Xiamen, China, 10–12 August 2018; pp. 1–5.
10. Liu, H.I.; Tseng, Y.W.; Chang, K.C.; Wang, P.J.; Shuai, H.H.; Cheng, W.H. A DeNoising FPN with Transformer R-CNN for Tiny Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–15. [\[CrossRef\]](#)
11. Zhou, Z.; Zhao, L.; Ji, K.; Kuang, G. A Domain Adaptive Few-Shot SAR Ship Detection Algorithm Driven by the Latent Similarity between Optical and SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–18. [\[CrossRef\]](#)
12. Zhang, Y.; Cai, Z. CE-RetinaNet: A channel enhancement method for infrared wildlife detection in UAV images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–12. [\[CrossRef\]](#)
13. Wang, Z.; Hou, G.; Xin, Z.; Liao, G.; Huang, P.; Tai, Y. Detection of SAR image multiscale ship targets in complex inshore scenes based on improved YOLOv5. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5804–5823. [\[CrossRef\]](#)
14. Bai, Z.; Pei, X.; Qiao, Z.; Wu, G.; Bai, Y. Improved YOLOv7 Target Detection Algorithm Based on UAV Aerial Photography. *Drones* **2024**, *8*, 104. [\[CrossRef\]](#)
15. Tang, Y.; Wang, T.; Liu, D. MFFAGAN: Generative Adversarial Network With Multi-Level Feature Fusion Attention Mechanism for Remote Sensing Image Super-Resolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6860–6874. [\[CrossRef\]](#)
16. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–9 June 2023; pp. 1–5.
17. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
18. Deng, W.; Li, X.; Xu, J.; Li, W.; Zhu, G.; Zhao, H. BFKD: Blockchain-based federated knowledge distillation for aviation Internet of Things. *IEEE Trans. Reliab.* **2024**. [\[CrossRef\]](#)
19. Li, X.; Zhao, H.; Xu, J.; Zhu, G.; Deng, W. APDPFL: Anti-poisoning attack decentralized privacy enhanced federated learning scheme for flight operation data sharing. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 19098–19109. [\[CrossRef\]](#)
20. Luan, T.; Zhou, S.; Liu, L.; Pan, W. Tiny-Object Detection Based on Optimized YOLO-CSQ for Accurate Drone Detection in Wildfire Scenarios. *Drones* **2024**, *8*, 454. [\[CrossRef\]](#)
21. Varga, L.A.; Kiefer, B.; Messmer, M.; Zell, A. Seadronessee: A maritime benchmark for detecting humans in open water. In Proceedings of the IEEE/CVF winter conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 2260–2270.
22. Gasienica-Jozkow, J.; Knapik, M.; Cyganek, B. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integr. Comput.-Aided Eng.* **2021**, *28*, 221–235. [\[CrossRef\]](#)
23. Cafarelli, D.; Ciampi, L.; Vadicamo, L.; Gennaro, C.; Berton, A.; Paterni, M.; Benvenuti, C.; Passera, M.; Falchi, F. MOBDrone: A drone video dataset for man overboard rescue. In Proceedings of the International Conference on Image Analysis and Processing, Lecce, Italy, 23–27 May 2022; pp. 633–644.
24. Han, J.; Yang, W.; Wang, Y.; Chen, L.; Luo, Z. Remote Sensing Teacher: Cross-Domain Detection Transformer with Learnable Frequency-Enhanced Feature Alignment in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [\[CrossRef\]](#)
25. Xu, H.; Ling, Z.; Yuan, X.; Wang, Y. A video object detector with Spatio-Temporal Attention Module for micro UAV detection. *Neurocomputing* **2024**, *597*, 127973. [\[CrossRef\]](#)
26. Li, J.; Zhang, Y.; Liu, H.; Guo, J.; Liu, L.; Gu, J.; Deng, L.; Li, S. A novel small object detection algorithm for UAVs based on YOLOv5. *Phys. Scr.* **2024**, *99*, 036001. [\[CrossRef\]](#)
27. Yi, H.; Liu, B.; Zhao, B.; Liu, E. Small object detection algorithm based on improved YOLOv8 for remote sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 1734–1747. [\[CrossRef\]](#)
28. Chen, C.; Zeng, W.; Zhang, X. HFPNet: Super Feature Aggregation Pyramid Network for Maritime Remote Sensing Small-Object Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 5973–5989. [\[CrossRef\]](#)
29. Xie, S.; Zhou, M.; Wang, C.; Huang, S. CSPPartial-YOLO: A lightweight YOLO-based method for typical objects detection in remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 388–399. [\[CrossRef\]](#)
30. Ge, W.; Yang, X.; Jiang, R.; Shao, W.; Zhang, L. CD-CTFM: A Lightweight CNN-Transformer Network for Remote Sensing Cloud Detection Fusing Multiscale Features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4538–4551. [\[CrossRef\]](#)

31. Chen, N.; Li, Y.; Yang, Z.; Lu, Z.; Wang, S.; Wang, J. LODNU: Lightweight object detection network in UAV vision. *J. Supercomput.* **2023**, *79*, 10117–10138. [[CrossRef](#)]
32. Pikun, W.; Ling, W.; Jiangxin, Q.; Jiashuai, D. Unmanned aerial vehicles object detection based on image haze removal under sea fog conditions. *IET Image Processing* **2022**, *16*, 2709–2721. [[CrossRef](#)]
33. Wang, C.; Shi, Z.; Meng, L.; Wang, J.; Wang, T.; Gao, Q.; Wang, E. Anti-Occlusion UAV Tracking Algorithm with a Low-Altitude Complex Background by Integrating Attention Mechanism. *Drones* **2022**, *6*, 149. [[CrossRef](#)]
34. Zeng, S.; Yang, W.; Jiao, Y.; Geng, L.; Chen, X. SCA-YOLO: A new small object detection model for UAV images. *Vis. Comput.* **2024**, *40*, 1787–1803. [[CrossRef](#)]
35. Li, W.; Liu, D.; Li, Y.; Hou, M.; Liu, J.; Zhao, Z.; Guo, A.; Zhao, H.; Deng, W. Fault diagnosis using variational autoencoder GAN and focal loss CNN under unbalanced data. *Struct. Health Monit.* **2024**. [[CrossRef](#)]
36. Li, J.; Wen, Y.; He, L. Scconv: Spatial and channel reconstruction convolution for feature redundancy. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 6153–6162.
37. Li, F.; Guo, M.; Su, R.; Wang, Y.; Wang, Y.; Xu, F. A multi-frame fusion video deraining neural network based on depth and luminance features. *Displays* **2024**, *85*, 102842. [[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.