

## Article

# A Framework and Method for Surface Floating Object Detection Based on 6G Networks

He Li <sup>1,2</sup> , Shuaipeng Yang <sup>1</sup>, Jinjiang Liu <sup>1,\*</sup>, Yang Yang <sup>2</sup>, Michel Kadoch <sup>3</sup> and Tianyang Liu <sup>4</sup>

<sup>1</sup> Henan Engineering Research Center of Intelligent Processing for Big Data of Digital Image, School of Computer Science and Technology, Nanyang Normal University, Nanyang 473061, China

<sup>2</sup> State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>3</sup> Synchromedia Laboratory, École de Technologie Supérieure, Université du Québec à Montréal, Montreal, QC H3C 1K3, Canada

<sup>4</sup> Sydney Business School, Shanghai University, Shanghai 200444, China

\* Correspondence: 20111041@nynu.edu.cn; Tel.: +86-1523-605-3001

**Abstract:** Water environment monitoring has always been an important method of water resource environmental protection. In practical applications, there are problems such as large water bodies, long monitoring periods, and large transmission and processing delays. Aiming at these problems, this paper proposes a framework and method for detecting floating objects on water based on the sixth-generation mobile network (6G). Using satellite remote sensing monitoring combined with ground-truth data, a regression model is established to invert various water parameters. Then, using chlorophyll as the main reference indicator, anomalies are detected, early warnings are given in a timely manner, and unmanned aerial vehicles (UAVs) are notified through 6G to detect targets in abnormal waters. The target detection method in this paper uses MobileNetV3 to replace the VGG16 network in the single-shot multi-box detector (SSD) to reduce the computational cost of the model and adapt to the computing resources of the UAV. The convolutional block attention module (CBAM) is adopted to enhance feature fusion. A small target data enhancement module is used to enhance the network identification capability in the training process, and the key-frame extraction module is applied to simplify the detection process. The network model is deployed in system-on-a-chip (SOC) using edge computing, the processing flow is optimized, and the image preprocessing module is added. Tested in an edge environment, the improved model has a 2.9% increase in detection accuracy and is 55% higher in detection speed compared with SSD. The experimental results show that this method can meet the real-time requirements of video surveillance target detection.



**Citation:** Li, H.; Yang, S.; Liu, J.; Yang, Y.; Kadoch, M.; Liu, T. A Framework and Method for Surface Floating Object Detection Based on 6G Networks. *Electronics* **2022**, *11*, 2939. <https://doi.org/10.3390/electronics11182939>

Academic Editor: Djurdj Budimir

Received: 7 August 2022

Accepted: 14 September 2022

Published: 16 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** remote sensing; target detection; edge computing; SSD; MobileNetV3; SOC

## 1. Introduction

With the development of industry and agriculture, large amounts of wastewater are discharged, and the ecological environment of watersheds is seriously damaged. Floating objects on the water [1], water pollution [2], and water eutrophication [3] are all important causes of water pollution, which not only pollute and damage water resources but also threaten human safety and health. The traditional monitoring method is to deploy devices in the water that make use of sensors to monitor and analyze the water quality. This method can accurately measure local water quality pollution, but it cannot carry out quick statistical analysis of large areas of water pollution. For visible floating objects on the water surface, monitoring is often accomplished by watching a screen. Although this method is simple, the coverage is limited, and it requires a lot of labor and material resources; thus, it is not very efficient.

In addition to traditional methods, an increasing number of technologies are being used to protect water resources. The use of remote sensing satellites [4] for monitoring

the pollution of target waters enables the acquisition of information over a large area in a short period. With the development of deep learning, a method for detecting floating objects based on a convolutional neural network has been proposed [5]. The convolutional neural network (CNN) can recognize and classify floating objects by extracting, training, and learning their features. In practice, target detection in streaming video involves huge computational effort and therefore has high hardware requirements. Edge computing [6] has been proposed as a new method for intelligent video surveillance, which can significantly reduce video-processing latency and ensure real-time performance. In comparison with traditional methods, these methods exhibit the following characteristics:

- Large coverage area: remote sensing technology can achieve comprehensive scanning and monitoring of target waters and areas, and coverage of inland waters can reach 100%.
- High detection accuracy: deep-learning-based target detection can achieve highly accurate detection and classification of floating objects on the water surface.
- Good real-time performance: edge computing computes, analyzes, and stores data near the data source to reduce redundant data transmission and meet the demand for real-time performance in practical application scenarios.

Remote sensing technology often needs to be combined with ground sampling and analysis data to invert the overall data of the watershed. Chlorophyll content is an important indicator for detecting water-surface algae. Deep-learning-based target detection can help in classifying floating objects. Different network models perform differently. In addition to pursuing higher detection accuracy and faster detection speed, many researchers have also investigated the balance between detection speed and accuracy. Edge computing is mainly performed at edge nodes, and the limited computing power limits the size and computation power of the detection model. Optimization can yield a smaller and more accurate network model. The 6G networks will be a fully connected domain with integrated terrestrial wireless and satellite communications. To realize the comprehensive monitoring of floating objects on the water surface, a new framework based on 6G is proposed for the warning and identification of such objects. Remote sensing technology is used to issue warnings regarding pollution of large water areas. A new method based on SSD-MobileNetV3 in edge computing is being used to monitor floating objects on small water areas. The CBAM, system-collaborative optimization, image preprocessing, and key-frame extraction were designed to reduce the interference of complex backgrounds and improve calculation speed, detection accuracy, and overall stability.

## 2. Related Works

Ecological monitoring is dynamic, large-scale, long-term work. With the rapid development of science and technology, remote sensing technology has been widely used in ecological environment monitoring. Using remote sensing for pollution monitoring is a new application for this technology. It can achieve rapid and large-scale monitoring of the ground environment, and is often used for water pollution monitoring [7] and water eutrophication monitoring [8]. Qun et al. [9] used remote sensing technology for detection of the water of Nansihu Lake. First, the relevant data were subjected to sensor, geometric, and atmospheric correction, and water extraction. Taking chlorophyll and suspended solids as important water quality indicators, the authors established a remote sensing inversion model of the water area. The inversion model was then combined with the conventional water detection method to invert the water quality parameters. The method was found to be faster, broader in scope, and more credible for water quality evaluation. Xiao et al. [10] proposed a random forest-based algorithm to distinguish *Ulva prolifera* and *Sargassum* from multispectral satellite images. Differential analysis was performed, mainly by capturing the spectra of *Ulva prolifera* and *Sargassum* using the GF-1 satellite sensor. The method can be used in marine waters with similar environments for phytoplankton traceability and competitive succession with high accuracy and stability. It provides reference values for identifying algae, monitoring water blooms, and providing early warning in inland waters.

Researchers have used the Gaussian mixture model (GMM), deep learning, and other methods to study the recognition and classification of floating objects on the water surface. In 2019, Jin et al. [11] proposed an improved GMM-based automatic segmentation method (IGASM) to detect floating objects on the water surface. The method first maps the GMM results onto the HSV color space and detects light and shadow using the light and shadow discriminant function. Then, floating objects on still water are segmented by the background update strategy combined with the graph cuts algorithm to optimize the segmentation results based on the spatial information of video images. The experimental results showed that the method could effectively eliminate the effects of light and shadow and water ripples. The improved background update strategy enabled better segmentation of floating objects on the water surface. In 2021, Zhang et al. [12] added an improved anchor refinement module to the convolutional layer of the RefineDet model. High-level semantic features could be extracted, and different levels of features could be fused to improve detection accuracy. The parameter settings of anchor points could be adjusted according to the scale and aspect ratio distribution, and the focus-loss function could be used to solve the foreground–background imbalance problem caused by too many anchor points. The experiments showed that the method could basically meet the requirements of real-time performance and precision. He et al. [13] proposed an improved YOLOv5 water surface floating object detection algorithm. The method suppressed overfitting in network training by introducing smoothing labels. The original topology was also used to enhance the feature extraction of floating objects and to reduce the number of parameters and computational effort. The loss function of the model was also optimized to improve speed and accuracy. The experiments showed that the strategy was feasible for detecting floating objects on the water surface.

The emergence of edge computing makes up for the shortcomings of cloud computing. As a result of the proliferation of mobile and IoT devices, large volumes of multimodal data (e.g., audio, picture, and video) of physical surroundings can be continuously sensed on the device side [14]. Taking intelligent video surveillance as an example, it requires 24 h of video data processing, computing and storage, which will impose higher requirements on the equipment environment. Cloud computing cannot meet the demand of network and computing costs, and edge computing has the advantages of low latency, low bandwidth, and low cost, and has been applied in various fields.

Sun et al. [15] proposed an edge computing-enabled mobile video processing system. Due to the limited resources of edge devices, they cannot deploy high-precision network inference models. The authors propose using mobile edge computing units with cameras and cloud nodes as edge cloud nodes through which video streams are processed. This method first preprocesses the video stream, then uploads the results to the upper node for processing, and utilizes the computing resources of the cloud to speed up data analysis. The experimental results show that this method can reduce video transmission delay and network overhead, provide a new idea for video processing in edge computing. Wang et al. [16] proposed an edge computing environment for accurate part model classification using a convolutional neural network-based element segmentation method. Xu et al. [17] proposed a cloud–edge collaboration framework for video surveillance in coal mines. The two are integrated, with cloud computing used for non-real-time and global tasks and edge computing used for real-time processing of local surveillance videos. A mixed edge-based and cloud-based framework with the final goal of PM2.5 value prediction is proposed in the literature [18]. In this scheme, the original and preprocessed data on a real-world dataset from air quality sensors distributed in Calgary, Canada, is used to evaluate the quality of predictions. The above methods effectively solve practical problems such as industrial production, environmental monitoring, and urban security through the collaboration of cloud computing and edge computing. The main work related to floating objects is summarized in Table 1.

**Table 1.** The main work related to floating objects.

| Method                                       | Main Technology                        | Advantages                                | Disadvantages                          |
|--|--|---|--|
| Water eutrophication monitoring [8]          | Remote sensing and MODIS               | Large coverage area                       | High hardware requirements             |
| Water quality evaluation [9]                 | Remote sensing and Sensors             | Wide monitoring range and high efficiency | High cost and low recognition accuracy |
| Monitoring water blooms [10]                 | Remote sensing and Random forest-based | high accuracy and stability               | High hardware requirements             |
| IGASM [11]                                   | GMM-based and image segmentation       | Identify and segment floating objects     | Unable to perform target detection     |
| Detection of Surface Floating Object [12]    | RefineDet-based                        | High accuracy                             | Low efficiency                         |
| Water surface floating object detection [13] | YOLOv5-based                           | Less parameters and calculation           | Low detection rate                     |
| mVideo [15]                                  | Edge computing video processing        | Low transmission delay                    | Lack of actual deployment              |
| Visual sorting System [16]                   | Cloud-Edge Computing and CNN-based     | high accuracy                             | High hardware requirements             |
| Surveillance Video Analysis System [17]      | Cloud-edge Computing and FL-YOLO       | Lower calculation smaller model scale     | High cost and low real-time            |

However, with the development of the chip industry, system-on-a-chip (SOC) [19] is widely used, which can provide richer computing resources for edge devices. In this paper, SOC and edge devices are combined as edge computing nodes to implement real-time floating object detection on water.

With the above-mentioned surface floating object detection methods, real-time video detection cannot be performed directly, and traditional cloud-based intelligent video surveillance has difficulty meeting the demand for real-time performance. The emergence of edge computing can help to change this situation. Running AI models at the edge requires not only improving the computational storage capacity of edge devices, but also optimizing network models to adapt to them. Compared to existing methods for detecting floating objects on the water, our proposed scheme has the following improvements:

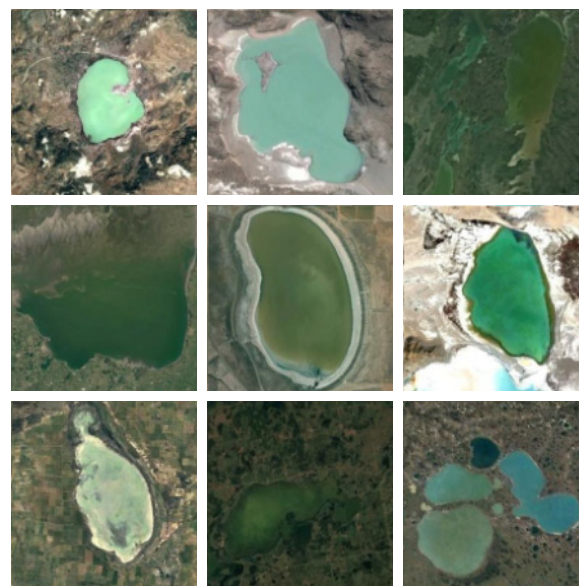
1. Remote sensing satellites can provide a larger range of monitoring data. We can monitor the chlorophyll content of water using remote sensing satellites and use that value to determine whether to conduct early warning.
2. After the early warning is issued, a UAV is used to perform aerial photography for detection of target waters, and the classification and detection of identifiable planktonic algae. The UAV and surveillance cameras around the water are the edge devices that generate data, combined with the SOC as the edge computing node.
3. The SSD network is deployed at edge nodes for floating object detection. Prior to that, we replaced VGG16 in the SSD network with MobileNet, reducing the computational cost to accommodate edge nodes.
4. By adding the key-frame extraction module, the frame difference method can effectively determine changes of floating objects on the water surface, and the detection of floating objects in key frames will help in capturing important information.
5. Image preprocessing is applied to key frames, including median filtering to remove noise, and Laplace sharpening, which can help the detection model to extract floating object features.

This paper is organized as follows: Section 3 describes the work related to water basin monitoring and early warning by satellite remote sensing; Section 4 presents the SSD-MobileNet network and the optimization improvement of the model; and Section 5 presents the deployment of the edge computing architecture and the analysis of the experimental results.

### 3. Framework of Remote Sensing Monitoring and Early Warning

The chlorophyll content in a water body is an important indicator to evaluate water quality and eutrophication. The reflection spectrum of a normal water body is mainly in the blue and green wavelengths, with a certain degree of absorption of other wavelengths, and the absorption capacity is the strongest in the near-infrared band. Due to the steep slope effect of chlorophyll in phytoplankton in visible and near-infrared wavelengths, an increase in chlorophyll concentration will weaken the absorption capacity of the water column. Estimating chloroplast concentration from the spectral reflectance of water provided by satellites is a commonly used monitoring method. We used this method to determine the planktonic biomass at the water surface.

The remote sensing monitoring and early warning process used in this paper included data collection [20] (shown in Figure 1), data preprocessing, water quality inversion, and abnormal warning.

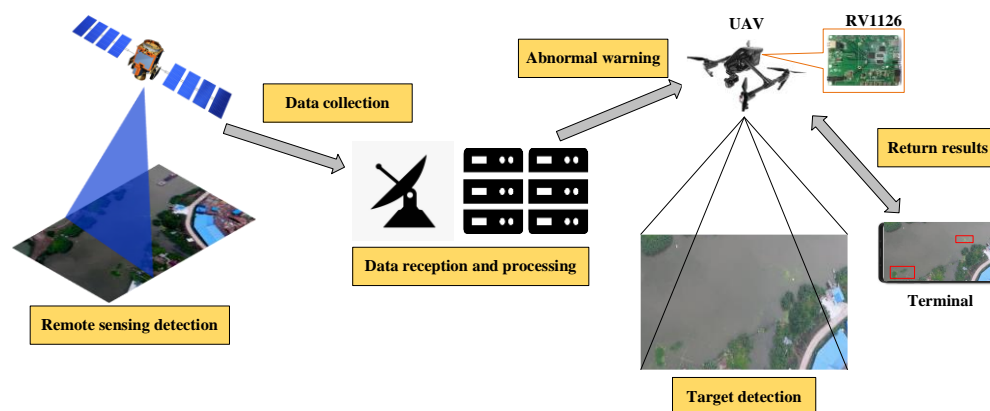


**Figure 1.** Satellite remote sensing data of related waters.

The process of remote sensing monitoring and early warning based on 6G is shown in Figure 2. First, the satellite collects global remote sensing image data and transmits the data to the ground receiving station through 6G. Several high-definition images in different wavelength bands of the water to be monitored can be acquired through ground receiving stations. Then, preprocessing procedures, such as radiometric calibration, atmospheric correction, multispectral correction, and image stitching, are performed [21]. The accuracy of radiation correction is an important indicator of the quality of satellite images. The China Resources Satellite Application Center provides absolute radiation correction coefficients, which can be used to calibrate GF-1 data and realize the conversion of DN values to radiation brightness values. The atmospheric correction module in the software is then used for atmospheric correction processing. Geometric correction then digitally performs a point-by-point fine correction of the image, and finally multiple images of the target water are stitched into a mosaic to ensure coverage of the whole water body. Remote sensing data and water quality analysis data of water are combined to invert the water quality parameters. When the detection result is abnormal, the ground station will send



the abnormal location information to the UAV through 6G. Then, the UAV will perform target area aerial photographic detection to detect the planktonic algae in the water with the real-time target.



**Figure 2.** Remote sensing monitoring process.

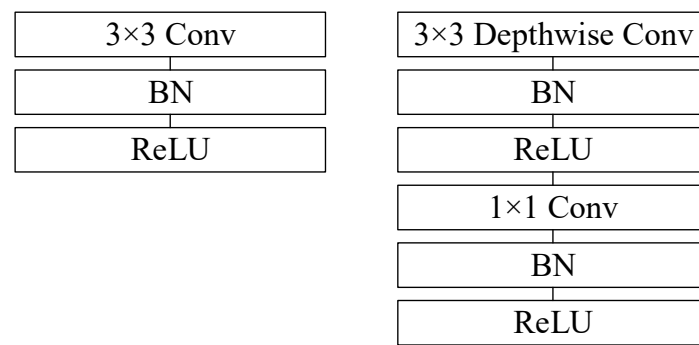
When the planktonic algae in the water are growing in large numbers, it leads to increased chlorophyll in the water column, and at this time the absorption of NIR wavelengths on the water surface is significantly weakened. In this paper, chlorophyll-a content is used as the main reference indicator. When the index is abnormal, the UAV target detection is carried out in the designated area according to the analysis results of remote sensing data. The UAV is equipped with the MobileNetV3-SSD target detection algorithm, which can detect and classify a variety of planktonic algae and other common floating debris on the target water surface. The detection data are uploaded in real time using the edge computing method to facilitate the next step.

## 4. Detection Algorithm

### 4.1. SSD-MobileNetV3

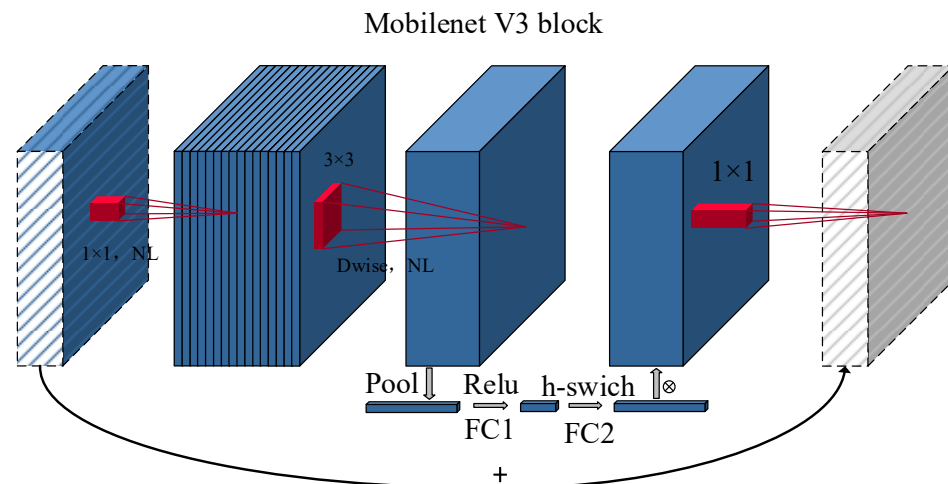
In general, there are two types of target detection methods: a two-stage algorithm based on proposed regions, such as R-CNN [22], and a single-stage algorithm based on regression, such as YOLO [23] and SSD [24]. In particular, SSD is a common single-layer object detection algorithm that uses the regression idea of YOLO to transform the object into a simple regression problem. SSD uses a pyramid feature layer-based detection method that performs both softmax classification and location regression on feature maps of different sizes. At the same time, SSD borrows the idea of anchor in Faster R-CNN, using different scales and aspect ratios and other prior frames, which will be more accurate for detecting and localizing objects of different sizes. The SSD network structure consists of a base network at the front end and an additional feature extraction layer at the back end. In the base network, the VGG16 network is used to extract basic features, and additional feature extraction layers extract more advanced features through a series of convolutional networks.

The VGG16 network has a large computational volume and parameters that limit deployment of the model in embedded systems. With the advent of MobileNet, it is entirely possible to replace standard convolutional convolution with deep separable convolution, which can reduce the large number of computations and parameters and effectively achieve model compression. Most models can be compressed in this way, which is very friendly for the deployment of models in embedded systems. As shown in Figure 3, MobileNet decomposes the standard convolution in the original network into a deep convolutional layer and a point convolutional layer, each followed by a batch normalization (BN) layer and a ReLU-activation function.



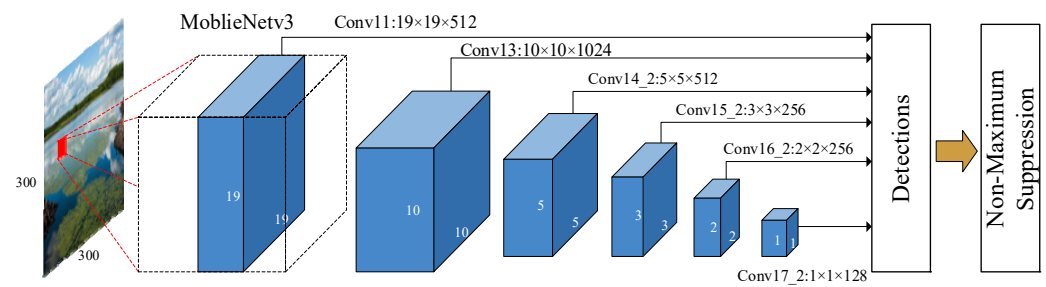
**Figure 3.** (Left) Standard convolution; (right) depth-wise separable convolution.

MobileNetV2 adds linear bottleneck and inverted residual to MobileNetV1. Linear bottleneck not only reduces the computational effort, but also effectively solves the problem of feature information loss caused by nonlinear activation layers. MobileNetV3 further adds the squeeze-and-excitation (SE) module. As shown in Figure 4, the channel weights are represented by the global pooling of each channel of the output feature matrix and then by two fully connected layer output vectors. The method allows the network to perform feature recalibration, where learning can automatically capture the importance of each feature channel, emphasizing important features and suppressing unimportant ones.



**Figure 4.** MobileNetV2 + squeeze-and-excite.

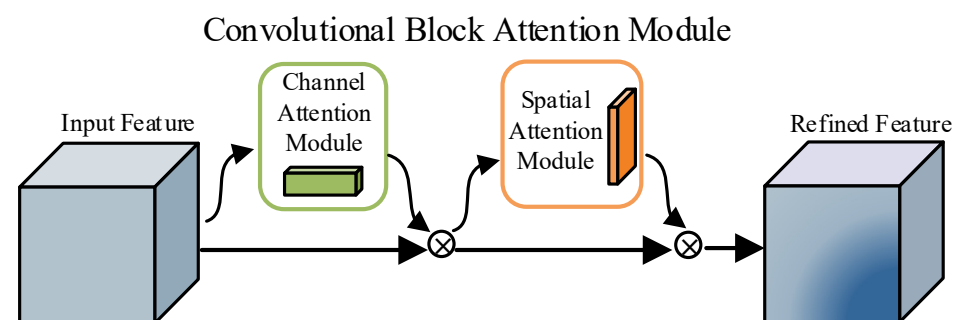
The network structure of SSD-MobileNetV3 [25] is shown in Figure 5. Obviously, MobileNetV3 replaces the original VGG16 network, and the whole network adopts multi-scale feature detection. The six scales of feature information point (the white numbers in Figure 5 are the dimensions of the feature map) to the detection module, which can realize judgment of the target location and category, and the next step is to filter out the redundant target boxes by the non-maximum suppression (NMS) algorithm.



**Figure 5.** Network structure of SSD-MobileNetV3.

#### 4.2. Convolutional Block Attention Module (CBAM)

The SE module in the MobileNet network is a channel attention mechanism. The SE module uses the same processing for the features in each channel. However, it is easy to ignore the information interactions in the space. As shown in Figure 6, CBAM [26] contains two modules, the channel attention and spatial attention modules. CBAM uses parallel pooling to apply both global average and global maximum pooling to the input feature maps, which can effectively reduce the loss of feature information. SAM applies global average and global maximum pooling to the input feature map, then stitches the two channels together and performs convolution and activation operations. This method is used to enhance specific target regions and weaken irrelevant background regions. In this paper, CBAM is used to replace the SE module. For the detection of floating objects on the water surface, this method can reduce the interference of complex backgrounds to a certain extent.



**Figure 6.** Overview of convolutional block attention module.

#### 4.3. Data Augmentation

During actual testing, we found that when the cameras were deployed farther away, more small targets were detected and were prone to false and missed detection. To address this problem, a small target data augmentation (STDA) module was added in the training process to increase the number of small targets in the samples and enhance the training of the network on small targets, in order to obtain better robustness.

Data augmentation allows for more adequate model samples and better generalization ability. Commonly used data augmentation methods include flipping, mirroring, and color-gamut transformation. As shown in Figure 7, the STDA method splices four randomly selected samples from the dataset, with each image scaled down to a random scale  $P$ . The scaled-down image is flipped, mirrored, and contrast-enhanced, etc., using random data enhancement. The scaled-down image is stitched together in the same way as the original, and then combined with the original image. Random box selection is performed in the combined images, and the boxed images are fed into the network as new samples for training.



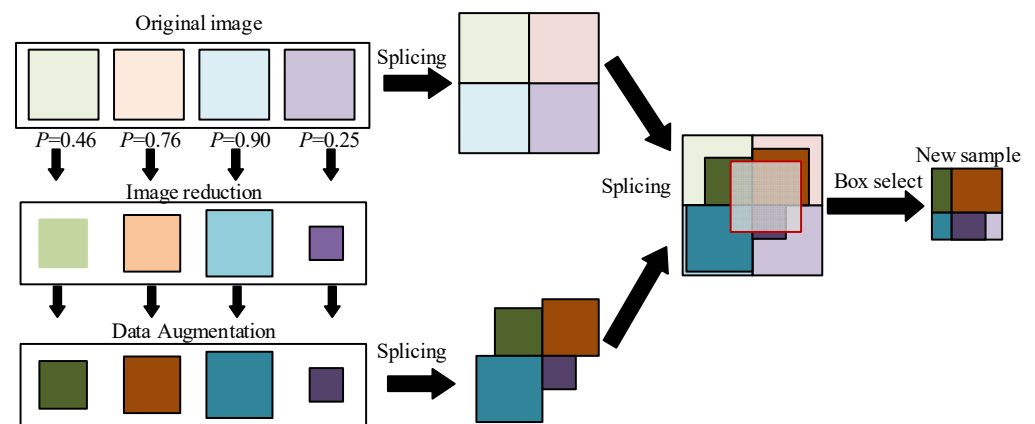


Figure 7. Small target data augmentation.

## 5. System Optimization

### 5.1. System Analysis

In the early stages of model design, we had to consider application scenarios and some of the constraints imposed by resource allocation. Complex models often require a large amount of computation, which is difficult to afford with the resource allocation of edge devices. Adapting the network model to edge devices and making it perform better is also an important task. As shown in Figure 8, a number of influencing factors are tuned and optimized according to the application requirements, resulting in faster calculations and more stable overall performance. In real-time detection, the amount of model computation is an important factor affecting the speed of detection. In this paper we replace the VGG16 network in SSD with MobileNet and quantify the model; this method can significantly reduce the amount of model computation. The SOC in the edge device is the main component of the whole system and the most critical unit for performing data-processing calculations. Adjusting the model parameters according to the hardware performance will enable the hardware to perform better.

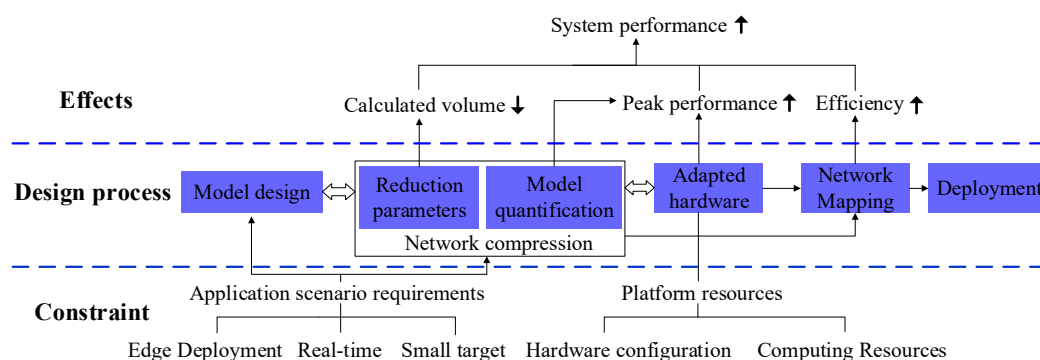
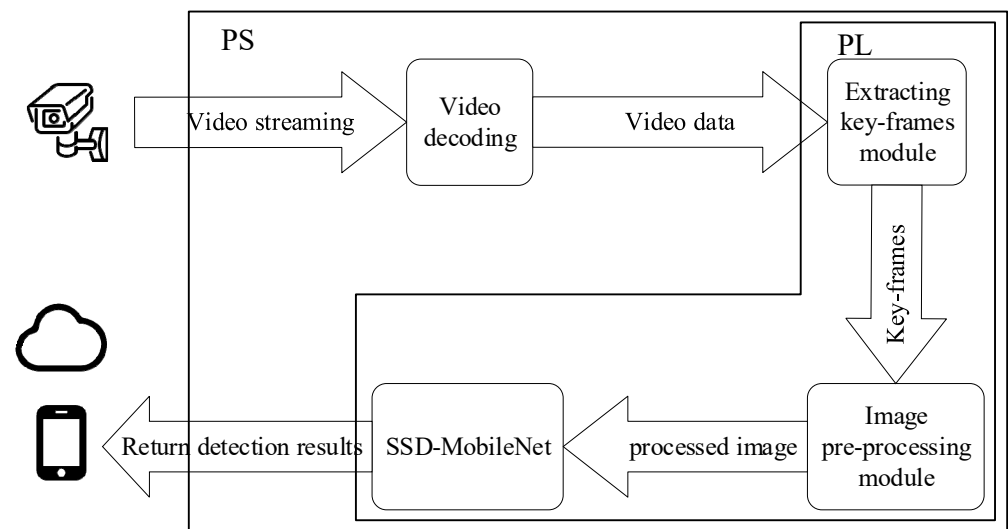


Figure 8. System design process.

In addition, in order to further simplify the process and improve the detection accuracy, we optimized and improved the processing system (PS). As shown in Figure 9, we added an extract key-frame module and an image-preprocessing module to the programmable logic (PL). The main function of the extraction key-frame module is to extract the moving video frames of floating objects from the video stream. These video frames contain important information for the detection of floating objects, and the processing of redundant data can be reduced by detecting these key frames. It will effectively reduce the amount of calculation. The image-preprocessing module is used to de-noise and sharpen key frames to further improve the detection accuracy. These two modules are described in detail in the following subsections.



**Figure 9.** Video detection process.

### 5.2. Extracting Key Frames

To further simplify the detection process, improve the speed of detection, and meet the demand for real-time performance in practical application scenarios for edge devices fixed on the shore, a key-frame extraction module was added to their detection process. In this paper, video frames that can reflect increase and decrease, and displacement changes of floating objects in the water, are used as key frames. The inter-frame difference method [27] can help us to quickly calculate and extract key frames. In this paper, the two-frame difference method is adopted to perform the difference operation between the  $n$ th and  $n - 1$ th frames of two temporally consecutive images, and the specific algorithm is as follows:

- Let  $A$  be the whole frame image, and the  $n$ th frame image and  $n - 1$ th frame image in the video sequence be  $f_n$  and  $f_{n-1}$ .
- The grayscale values of the corresponding pixel points of the two frames are denoted as  $f_n(x, y)$  and  $f_{n-1}(x, y)$ . Then, the absolute value of the difference between the grayscale values of the corresponding pixel points in the two frames is summed. The calculation process is given in Equation (1):

$$D_n = \text{SUM}|f_n(x, y) - f_{n-1}(x, y)|, (x, y) \in A \quad (1)$$

when  $D_n$  exceeds a certain threshold, it is determined that there is a floating object moving in this video frame, and the frame is used as a key frame. A threshold value that is too small cannot suppress many noise points in the image, and a threshold value that is too large tends to obscure the target information. Fixed thresholds cannot adapt to light changes in the scene. In this paper, we added an addendum to the determination condition to adjust the threshold value according to the overall lighting. The key-frame determination conditions are given in Equation (2):

$$D_n = T + \lambda \frac{1}{N_A} \sum_{(x, y)} |f_n(x, y) - f_{n-1}(x, y)| \quad (2)$$

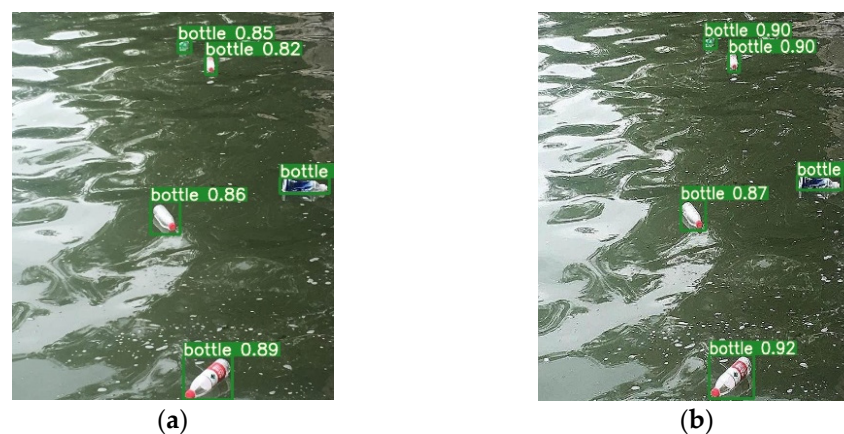
where  $N_A$  is the total number of pixels in the area to be detected,  $\lambda$  is the rejection factor for illumination, and  $A$  is the whole image. The addition term indicates the change in illumination in the whole image.

If the change in illumination in the scene is small, the value of this term tends to be zero. If the change in illumination in the scene is significant, the value of this term increases significantly, and the right-hand side of the judgment condition increases adaptively, thus effectively suppressing the effect of light changes on the detection results of moving targets.

### 5.3. Image Preprocessing

The water-surface environment is complex and easily disturbed by other factors in the process of floating object detection, resulting in loss of detection accuracy and even false detection or omission. In this paper, an image-preprocessing module was added before the model detection for median filter noise elimination and Laplacian sharpening of key frames, to preserve floating object edge information while eliminating noise. Passing the processed image into the model detection is beneficial to the feature extraction of the image by the model, which can effectively improve the detection accuracy.

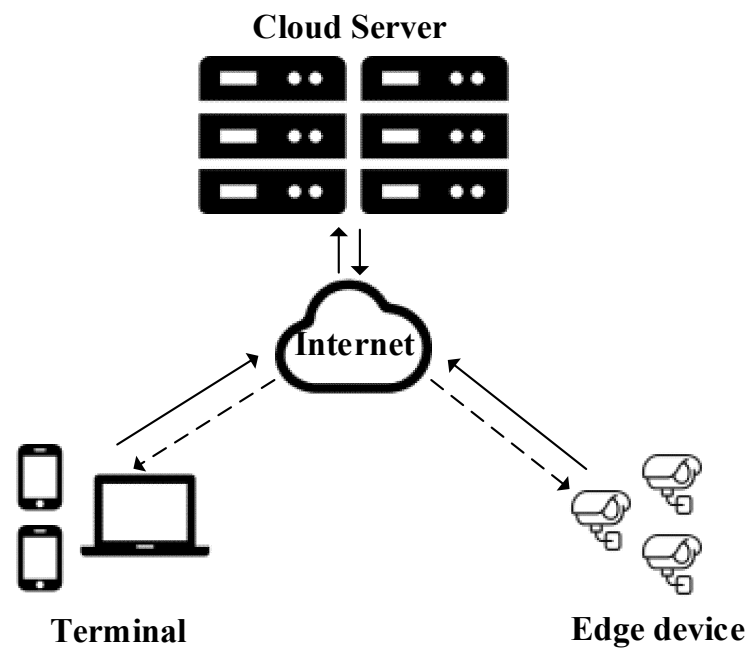
Median filtering is a nonlinear signal-processing method, so it is a nonlinear filter and a statistical-sorting filter. First, we specify the sliding window size, take the median of the grayscale values of the neighboring pixels in the center of the window, and replace the value of the center pixel with the calculated median value. The key frames are de-noised using median filtering, which can effectively suppress the noise effect and keep the edge effects of the image without making it too blurry. The image is then further processed with Laplacian sharpening. When the grayscale value of the central pixel is lower than the average grayscale of other pixels in its neighborhood, the grayscale of the central pixel will be further reduced. When the grayscale value of the central pixel is higher than the average value of other pixels in its neighborhood, the grayscale of this pixel should be further improved. By sharpening the image in this way, the details can be enhanced, and the edges can be highlighted. As shown in Figure 10, the detection precision of this method is slightly improved compared with the original model.



**Figure 10.** (a) Detection result of original image; (b) detection result of processed image.

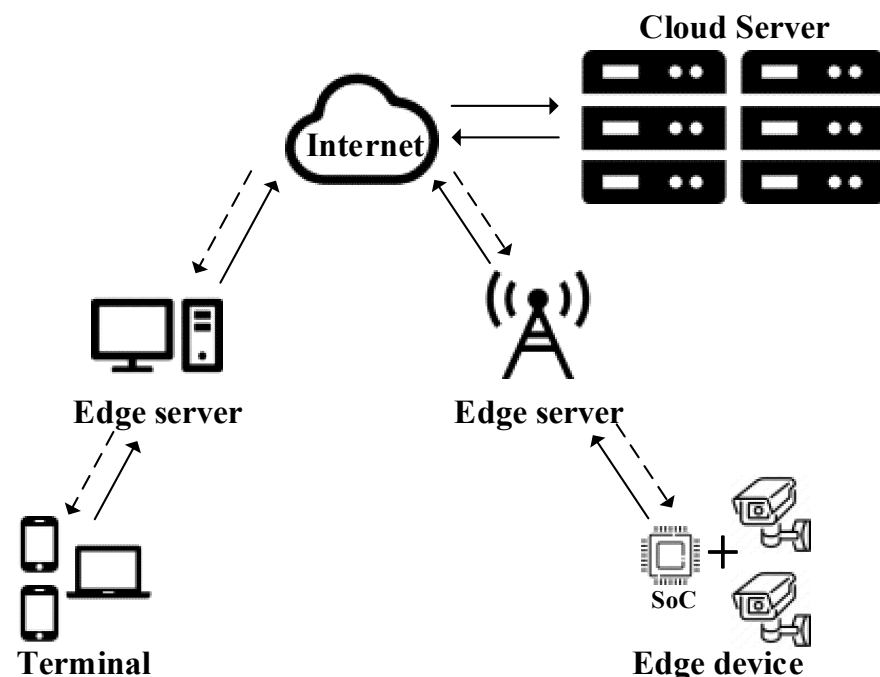
### 5.4. Edge Deployment

The advent of convolutional neural network-based target detection has rapidly moved intelligent video analysis from theory to practical application. Deep convolutional neural networks require large amounts of computation and must rely on hardware such as a graphics processing unit (GPU) to achieve this. The traditional cloud-based real-time video streaming analysis model is shown in Figure 11. The video data are transmitted to the cloud server in the network center in real time through the Internet, and the data are cleaned, stored, analyzed, and reasoned by the cloud server, then the reasoning results are returned to the terminal device. This model has a stable overall structure and is widely used in various business scenarios. However, problems such as large bandwidth consumption, high-transmission delay, unreliable network, and difficult privacy protection still need to be solved.



**Figure 11.** Traditional video stream analysis architecture based on cloud computing.

The emergence of SOC has provided arithmetic support for edge computing, making its deployment and popularity possible. In this paper, we apply the model of edge computing to intelligent video surveillance by sinking the cloud server at the center of the network to an edge node that is physically close to the video source. SOC with some computational power is embedded in the camera as edge nodes, and the above detection model is deployed. The camera transmits the captured video stream data to the SOC, which then decodes it according to the frame rate, resolution, and other parameters, and the video coding protocol. The above key-frame extraction algorithm is then used to extract key frames from the video stream and pass them into the model for target detection. The edge analysis architecture [28] is shown in Figure 12.



**Figure 12.** Edge computing-based video streaming analytics architecture.

### 5.5. Limitations of the Method

The calculation of edge nodes used in edge computing platforms mainly depends on SOC, which has limited computing power. In this paper, the channel pruning of the trained network model is carried out to reduce the computational burden and basically achieve the demand of edge adaptation. However, in the actual deployment of the network model, there will be some phenomena such as the inability to detect the target continuously, missing detection, and the occasional jump and drift of the detection box. Therefore, it is an important challenge to reduce the computational burden of the model and reduce the accuracy loss while ensuring the real-time performance of object detection. In the future, we will use filtering and smoothing methods to predict targets and reduce missed detection. Meanwhile, multi-thread optimization, inter-frame optimization, and algorithm co-optimization will be used to shorten the processing delay.

In addition, small object detection is always a difficult problem in the field of object detection. In the target detection task, convolutional neural network achieves localization and classification by extracting the feature information of the target. Obviously, the amount of feature information carried by the target directly affects the final prediction result. Small objects occupy a low proportion of pixels in the image and carry less effective feature information, which makes the detection and recognition of small objects more difficult. In the water environment, special background factors such as light, ripples, and reflections have to be considered, which can lead to a false detection of the results. Meanwhile, in practical application, the different types of floating objects on the water surface are various, and the size distribution is different, which also brings great challenges to the identification and detection. Aiming at the above problems, a data enhancement method is used for small targets to increase the number of small target samples and improve the generalization ability of the model. The attention mechanism is added to make the network pay more attention to the key information carried by small targets. Experiments show that these methods can improve the accuracy of small target detection. However, there are only four types of floating objects in the dataset used in this paper. The objects with a small data amount and insignificant features are not included in the dataset due to the difficulty of collection. In the future, we will increase the collection and sorting of such image data and improve the floating object dataset. The multi-size detection strategy and cross-feature layer-fusion method will be used to improve the accuracy of small object detection.

## 6. Experimental Analysis

### 6.1. Datasets

In this paper, the experimental data on river floaters were mainly obtained from publicly available datasets, such as ImageNet [29] and COCO [30], manual photography, and relevant images using web crawler techniques. Then, LabellImg software was used to label the images, and a dataset of floating objects on the water surface was produced in VOC format. The dataset consisted of four main categories: bottles, plastic bags, planktonic algae, and dead fish. The dataset was expanded by rotation, contrast enhancement, and mirroring, as shown in Figure 13. A total of 22,000 images were collected from the dataset, and the statistics are shown in Table 2.

**Table 2.** Float dataset.

| Classes          | Training Set | Validation Set | Test Set |
|------------------|--------------|----------------|----------|
| Bottles          | 8000         | 2700           | 2300     |
| Plastic bags     | 6000         | 2200           | 1800     |
| Planktonic algae | 6000         | 3200           | 2800     |
| Dead fish        | 2000         | 1000           | 1000     |





**Figure 13.** Dataset types and expansion effects. Left to right: original images, increased-contrast images, and mirror images.

## 6.2. Experimental Environment

The experimental environment was divided into model-training and edge deployment environments. The model-training software environment was Ubuntu 18.04, python 3.8.8, using the Pytorch framework. The model-training hardware environment was GeForce RTX 3090, Intel(R) Xeon(R) CPU E5-2678 v3. The edge deployment environment used a 2 megapixel 1/1.8-inch (charge coupled device, CCD) CMOS smart capture camera and a RV1126 chip.

## 6.3. Experimental Results and Analysis

The metrics generated during the training of the network model were the criteria for evaluating the quality of the model and provide an objective picture of the model's performance. For the performance evaluation, we selected accuracy rate  $P$ , recall rate  $R$ , average accuracy rate  $AP$ , and detection speed  $FPS$  to represent the performance of the model.  $P$  and  $R$  are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

where  $TP$  indicates the number of correctly detected floating objects,  $FP$  indicates the number of non-floating objects, and  $FN$  indicates the number of undetected floaters.

In this paper, we replaced the SE module in the network with CBAM before training SSD-MobileNetV3 and added the small target data augmentation (STDA) module during network training. An ablation study was conducted for the above improvements and the experimental results are shown in Table 3.

**Table 3.** Ablation study of detection precision of test set.

| Convolutional Block Attention Module | Small Target Data Augmentation | Accuracy Rate (%) | Recall Rate (%) | Average Accuracy Rate (%) |
|--------------------------------------|--------------------------------|-------------------|-----------------|---------------------------|
| ✓                                    | ✓                              | 91.26             | 88.31           | 87.34                     |
|                                      |                                | 93.10             | 91.35           | 89.66                     |
| ✓                                    | ✓                              | 93.52             | 90.16           | 91.50                     |
|                                      |                                | 94.31             | 91.32           | 92.17                     |

Where  $\checkmark$  indicates that the CBAM or STDA has been added. As can be seen from the experimental data, the values of the indicators are significantly lower when no improvements are made to the model. When CBAM and STDA were added,  $P$  improved by 2.01 and 2.48%,  $R$  improved by 3.44 and 2.10%, and  $AP$  improved by 2.66 and 4.76%, respectively. When both modules were added at the same time, the three metrics improved by 3.34, 3.41, and 5.53% respectively. This shows that including CBAM in the network and using the proposed STDA method in this paper can effectively improve the model detection accuracy.

To verify the effectiveness of the improvements, we deployed SSD, SSD-MobileNetV3, and the improved methods on edge devices for experiments. In all, 1000 data items were used as the validation dataset, including 800 small target and 200 regular data items. The evaluation metrics included  $P$ , mean average precision (mAP) (0.5), mAP (0.75), and frames per second (FPS).  $P$  is the detection accuracy; the two evaluation metrics, mAP (0.5) and mAP (0.75), are set according to different intersection over union (IOU) thresholds; and FPS is the number of image frames per second detected.

Through edge-testing experiments, SSD still maintained high detection precision, and SSD-MobileNetV3 replaced the computationally intensive VGG16 network with significantly higher detection speed. In Table 4, the addition of the extraction key-frame module and image-preprocessing module introduced additional computational effort into the system but showed good results in terms of speed and accuracy. The experimental data show that our method improved detection accuracy by 2.9% and 5.5% compared to the other two methods, and detection speed by 55% compared to SSD. A detection speed of 33 frames per second is perfectly suited to real-time requirements at the edge.

**Table 4.** Results of different networks tested at the edge.

| Method          | Precision | mAP (0.5) | mAP (0.75) | FPS |
|-----------------|-----------|-----------|------------|-----|
| SSD             | 92.61     | 95.38     | 90.32      | 22  |
| SSD-MobileNetV3 | 90.36     | 90.53     | 88.29      | 36  |
| Our Method      | 95.32     | 97.69     | 93.34      | 34  |

## 7. Conclusions

In this paper, a video-monitoring system for floating objects on the water surface was improved by combining it with satellite remote sensing data to monitor the water surface, covering a larger area and making the data professional and diverse. Improving the target detection model based on edge computing allowed the model to meet real-time performance requirements and have high detection accuracy. With drones and fixed monitors as the main detection methods, a multidirectional and more three-dimensional monitoring mechanism was established. The experimental results show that the improved detection system can meet real-time performance requirements and improve detection accuracy, and it has better detection accuracy for small targets. Therefore, the method in this paper can meet the detection requirements of embedded mobile terminals and provide a feasible technical solution for embedded edge computing.

**Author Contributions:** Conceptualization, H.L. and S.Y.; methodology, H.L.; software, S.Y.; validation, formal analysis, and investigation, H.L. and J.L.; resources, T.L. and J.L.; data curation, M.K. and Y.Y.; writing—original draft preparation, H.L. and S.Y.; writing—review and editing, visualization, supervision, project administration, and funding acquisition, M.K. and Y.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partly supported by the National Natural Science Foundation of China (grant no. 62002180), the Natural Science Foundation of Henan Province (grant no. 202300410301), the General Project of Humanities and Social Sciences Research of Henan Institutions of Higher Learning (grant no. 2021-ZZJH-262), the Scientific and Technological Project in Henan Province of China (grant nos. 212102310481, 222102320369, 212102210169), and the Key Scientific Research Projects of Colleges and Universities in Henan Province (grant nos. 21A520033, 22A520037).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Zou, S.H.; Li, P.; Fang, N.S. Real-time detection method of surface floating objects based on deep learning. In Proceedings of the 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Xuzhou, China, 16–19 October 2020; pp. 174–177.
2. Zhang, K.; Amineh, R.K.; Dong, Z.; Nadler, D. Microwave Sensing of Water Quality. *IEEE Access* **2019**, *7*, 69481–69493. [CrossRef]
3. Zhou, B.T.; Zhang, Y.Y.; Shi, K. Research progress on remote sensing assessment of lake nutrient status and retrieval algorithms of characteristic parameters. *Natl. Remote Sens. Bull.* **2022**, *26*, 77–91.
4. Chu, H.Q.; Huang, S.; Wang, X.F. Monitoring Water Quality Parameters of Taihu Lake Based on Remote Sensing Images and LSTM-RNN. *IEEE Access* **2020**, *8*, 188068–188081.
5. Zailan, N.A.; Mohd Khairuddin, A.S.; Khairuddin, U.; Taguchi, A. YOLO-based Network Fusion for Riverine Floating Debris Monitoring System. In Proceedings of the International Conference on Electrical, Communication, and Computer Engineering, Kuala Lumpur, Malaysia, 12–13 June 2021; pp. 1–5.
6. De Donno, M.; Tange, K.; Dragoni, N. Foundations and Evolution of Modern Computing Paradigms: Cloud, IoT, Edge, and Fog. *IEEE Access* **2019**, *7*, 150936–150948. [CrossRef]
7. Wang, J.Z.; Zhang, J.P.; Li, T.; Wang, X. Water Quality Analysis of Remote Sensing Images Based on Inversion Model. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4861–4864.
8. El-Alem, A.; Chokmani, K.; Laurion, I.; El-Adlouni, S.E.; Raymond, S.; Ratté-Fortin, C. Ensemble-Based Systems to Monitor Algal Bloom with Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7955–7971. [CrossRef]
9. Miao, Q.; Tan, X.H.; Liu, Z.Q.; Liu, C.; Li, Q.Q. Monitoring Chlorophyll-a and Suspended Substance in Nansi Lake, China through Remote Sensing Technology. In Proceedings of the 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, Shanghai, China, 21–22 December 2008; Volume 2, pp. 348–351.
10. Xiao, Y.F.; Liu, R.J.; Kim, K.; Zhang, J.; Cui, T.W. A Random Forest-Based Algorithm to Distinguish *Ulva prolifera* and *Sargassum* From Multispectral Satellite Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
11. Jin, X.L.; Niu, P.W.; Liu, L.F. A GMM-Based Segmentation Method for the Detection of Water Surface Floats. *IEEE Access* **2019**, *7*, 119018–119025. [CrossRef]
12. Zhang, L.L.; Wei, Y.X.; Wang, H.B.; Shao, Y.H.; Shen, J. Real-Time Detection of River Surface Floating Object Based on Improved RefineDet. *IEEE Access* **2021**, *9*, 81147–81160. [CrossRef]
13. He, X.Q.; Wang, J.C.; Chen, C.B.; Yang, X.Q. Detection of the Floating Objects on the Water Surface Based on Improved YOLOv5. In Proceedings of the 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 17–19 December 2021; Volume 2, pp. 772–777.
14. Zhou, Z.; Chen, X.; Li, E.; Zeng, L.K.; Luo, K.; Zhang, J.S. Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. *Proc. IEEE* **2019**, *107*, 1738–1762. [CrossRef]
15. Sun, H.; Yu, Y.; Sha, K.W.; Lou, B.D. mVideo: Edge Computing Based Mobile Video Processing Systems. *IEEE Access* **2020**, *8*, 11615–11623. [CrossRef]
16. Wang, Y.B.; Hong, K.J.; Zou, J.; Peng, T.; Yang, H.Y. A CNN-Based Visual Sorting System with Cloud-Edge Computing for Flexible Manufacturing Systems. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4726–4735. [CrossRef]
17. Xu, Z.; Li, J.Z.; Zhang, M. A Surveillance Video Real-Time Analysis System Based on Edge-Cloud and FL-YOLO Cooperation in Coal Mine. *IEEE Access* **2021**, *9*, 68482–68497. [CrossRef]
18. Soroush, O.; Francesco, C.; Terracina, G.; Liang, S.H.L. Enhanced air quality prediction by edge-based spatiotemporal data preprocessing. *Comput. Electr. Eng.* **2021**, *96*, 107572.
19. Sun, M.; Cao, Y.J.; Chiang, P.Y. Energy-aware Retinaface: A Power Efficient Edge-Computing SOC for Face Detector in 40nm. In Proceedings of the 2021 IEEE 14th International Conference on ASIC (ASICON), Kunming, China, 26–29 October 2021; pp. 1–4.
20. Gong, C.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. Available online: <http://www.escience.cn/people/JunweiHan/NWPU-RESISC45.html> (accessed on 28 July 2022).
21. Cao, Y.; Ye, Y.T.; Zhao, H.L.; Jiang, Y.Z.; Dong, J.P.; Yan, D.M. Analysis of the application potential of HMS-6 imagery for chlorophyll-a inversion in inland water bodies. *J. Geo-Inf. Sci.* **2022**, *24*, 546–557.
22. Zhou, Z.J.; Zhang, M.; Chen, J.F.; Wu, X.Q. Detection and Classification of Multi-Magnetic Targets Using Mask-RCNN. *IEEE Access* **2020**, *8*, 187202–187207. [CrossRef]
23. Luo, S.; Yu, J.; Xi, Y.J.; Liao, X. Aircraft Target Detection in Remote Sensing Images Based on Improved YOLOv5. *IEEE Access* **2022**, *10*, 5184–5192. [CrossRef]

24. Lyu, Z.W.; Jin, H.F.; Zhen, T.; Sun, F.Y.; Xu, H. Small Object Recognition Algorithm of Grain Pests Based on SSD Feature Fusion. *IEEE Access* **2021**, *9*, 43202–43213. [[CrossRef](#)]
25. Yu, G.Y.; Wang, L.; Hou, M.X.; Liang, Y.C.; He, T.H. An adaptive dead fish detection approach using SSD-MobileNet. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 1973–1979.
26. Luo, Y.N.; Wang, Z.S. An Improved ResNet Algorithm Based On CBAM. In Proceedings of the 2021 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 24–26 September 2021; pp. 121–125.
27. Li, W.J.; Yao, J.G.; Dong, T.Z.; Li, H.F.; He, X.J. Moving vehicle detection based on an improved interframe difference and a Gaussian model. In Proceedings of the 2015 8th International Congress on Image and Signal Processing (CISP), Shenyang, China, 14–16 October 2015; pp. 969–973.
28. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
29. Puri, D. COCO Dataset Stuff Segmentation Challenge. In Proceedings of the 2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 19–21 September 2019; pp. 1–5.
30. Chen, J.S.; Ran, X.K. Deep Learning with Edge Computing: A Review. *Proc. IEEE* **2019**, *107*, 1655–1674. [[CrossRef](#)]