



Article An Intelligent Detection Method for Small and Weak Objects in Space

Yuman Yuan ¹^[1], Hongyang Bai ¹,*^[1], Panfeng Wu ^{2,3}, Hongwei Guo ¹, Tianyu Deng ¹ and Weiwei Qin ⁴

- ¹ School of Energy and Power Engineering, Nanjing University of Science and Technology,
- Nanjing 210094, China; yym@njust.edu.cn (Y.Y.); njustghw@njust.edu.cn (H.G.); dty@njust.edu.cn (T.D.)
- ² Harbin Institute of Technology, Harbin 150006, China; 21b918144@stu.hit.edu.cn
 ³ Shandang Assesses Electronics Technology Basesses Institute Ventri 24470, China; 21b918144
- ³ Shandong Aerospace Electronics Technology Research Institute, Yantai 264670, China
 ⁴ Xi'an Pasaarch Institute of High Tech Xi'an 710025 China: guyu1982@pudt edu on
- ⁴ Xi'an Research Institute of High-Tech, Xi'an 710025, China; qww1982@nudt.edu.cn
- Correspondence: hongyang@njust.edu.cn

Abstract: In the case of a boom in space resource development, space debris will increase dramatically and cause serious problems for the spacecraft in orbit. To address this problem, a novel context sensing-YOLOv5 (CS-YOLOv5) is proposed for small and weak space object detection, which could realize the extraction of local context information and the enhancement and fusion of spatial information. To enhance the expression ability of feature information and the identification ability of the network, we propose the cross-layer context fusion module (CCFM) through multiple branches in parallel to learn the context information of different scales. At the same time, to map the small-scale features sequentially to the features of the previous layer, we design the adaptive weighting module (AWM) to assist the CCFM in further enhancing the expression of features. Additionally, to solve the problem that the spatial information of small objects is easily lost, we designed the spatial information enhancement module (SIEM) to adaptively learn the weak spatial information of small objects that need to be protected. To further enhance the generalization ability of CS-YOLOv5, we propose a contrast mosaic data augmentation to enrich the diversity of the sample. Extensive experiments are conducted on self-built datasets, which strongly prove the effectiveness of our method in space object detection.

Keywords: space object; object detection; feature fusion; data augmentation; convolutional neural networks (CNNs)

1. Introduction

With the continuous development of science and technology, the exploration of the space environment is gradually deepening. Satellites are widely used as infrastructure for modern communications, navigation, surveillance, and reconnaissance. Therefore, the number of satellites in space has increased dramatically, and a large quantity of space debris and abandoned satellites have posed a serious threat to the security and stability of the space environment. Figure 1 shows the growth curve of space debris quantities over time and the total mass of all space objects in Earth orbit statistically exceeding 10,400 tons [1]. Therefore, the space object monitoring system has gradually become a hot direction in many countries. The application of traditional telescopes in accurate positioning and continuous tracking of natural celestial bodies such as stars and planets has been very mature [2], but the movement of near-earth space objects such as artificial satellites, debris, and meteorites are different from natural celestial bodies, which makes the traditional space monitoring system unable to achieve effective performance. Currently, as an important information source of space situational awareness, image sensor provides a solid data foundation for space-based optical detection. It has intuitive, all-weather, and all-directional advantages, such as ORS-5 and SBSS-1 of the United States. Therefore, space-based optical detection system has been widely used in monitoring abandoned satellite debris, space threat object



Citation: Yuan, Y.; Bai, H.; Wu, P.; Guo, H.; Deng, T.; Qin, W. An Intelligent Detection Method for Small and Weak Objects in Space. *Remote Sens.* 2023, *15*, 3169. https:// doi.org/10.3390/rs15123169

Academic Editors: Yang Yang, Sanat K. Biswas and Xiaofeng Wu

Received: 30 March 2023 Revised: 13 June 2023 Accepted: 14 June 2023 Published: 18 June 2023



Copyright: © 2023 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/).



warning, space collision avoidance, satellite integrity monitoring, space debris warning, and so on.

Monthly Number of Objects in Earth Orbit by Object Type

Figure 1. Monthly number of objects in earth orbit by object type.

Traditional methods for object detection primarily focus on feature extraction (enhancing feature expression and resistance to distortion) and feature classification (improving classification accuracy and speed). Consequently, researchers have proposed various forms of features and classifiers, including SIFT [3], Hough [4], AdaBoost [5], DPM [6], etc. However, traditional object detection methods using hard-crafted features suffer from the following three drawbacks: (a) the designed features are low-level features and insufficient in expressing objects adequately; (b) the designed features are poorly distinguishable, leading to high classification error rates; (c) the designed features are specific, making it challenging to select a single feature for detection of multiple types of targets in various complex scenarios. Convolution neural networks (CNNs) have gradually emerged and been applied in various object detection fields, resulting in the mainstream method of current object detection. Therefore, it is of great significance to research space object detection based on deep learning and provide spacecraft and satellites with the ability to autonomously identify surrounding objects to avoid space collisions.

As mentioned above, CNNs have been widely used in various fields. However, there have been relatively few studies on space object detection and many deficiencies. Firstly, space-based optical imaging equipment captures a wide field of view with long imaging distances, resulting in small-sized objects in the images. The features of these small objects will gradually diminish during multiple downsampling operations performed by the network. Secondly, space images are affected by stray light from the Earth, resulting in low contrast of objects in backlight environments. The feature of weak objects may be submerged by the background features during the detection process. Thirdly, there may be a large-scale difference between different scale objects in space images, accompanied by dense phenomena. It becomes challenging for the network to distinguish the features of densely packed and small objects, leading to false and missed detection. Some typical examples of space images are shown in Figure 2. In addition to the aforementioned difficulties, there



is a scarcity of spatial data samples available for collection, resulting in insufficient prior information for the network to learn, which further affects detection accuracy.

Figure 2. Some typical examples of space images. (**a**) Small objects; (**b**) backlight environment; (**c**) multiple scales; (**d**) dense distribution.

To solve the above problems, we propose the context sensing-YOLOv5 (CS-YOLOv5) based on YOLOv5 [7] for space object detection. This network incorporates a cross-layer context fusion module, adaptive weighting module, and spatial information enhancement module to aid in the effective detection of weak and small objects in space. CS-YOLOv5 overcomes the difficulties of the small-scale and low contrast of objects in space object detection, enabling the extraction of local context information and the enhancement and fusion of spatial information. To tackle the problem of small-sized objects and objects with large-scale changes in the spatial images, we propose cross-layer context fusion modules (CCFM). By utilizing multiple parallel branches, the feature map is convoluted at different scales to learn the context information of different scales, thereby enhancing the feature representation of small objects. Within the CCFM, we incorporate an attention mechanism to propose an adaptive weighting module (AWM), which maps the smallscale features to upper-layer features, enhances the expression of effective information, suppresses the interference of useless information to the object features, and enhances the features of different scales. Aiming at the problem that the spatial information of small objects is easily lost, we propose a spatial information enhancement module (SIEM), which comprehensively learns the relative spatial relationships in different channels and orientations. To further improve the detection and generalization ability of CS-YOLOv5 for space objects, we propose a data augmentation method called contrast mosaic, which enhances the diversity and complexity of data while avoiding overfitting.

Due to the high cost of obtaining spatial image samples and the limited availability of datasets, synthetic datasets are currently the default approach for deep learning methods in space object detection tasks. For example, SPEED+ is the first dataset for vision-only spacecraft pose estimation and relative navigation, proposed by Stanford Uni Simone D'Amico's group [8]. It addresses the domain gap between synthetic training images and Hardware-In-the-Loop (HIL) test images. However, the images in this dataset primarily consist of single objects in backlight scenes. In addition to studying low-contrast objects,

this paper emphasizes densely distributed small objects. To meet the research needs, we constructed the near-earth space object (NSO) dataset, in which three types of objects, including debris, satellites, and meteorolites, are combined with the real Earth and starry sky as backgrounds. We simulate various object attitudes in space to ensure that the dataset samples closely resemble real scenes. The NSO dataset consists of 233,687 instances across 10,574 images. Extensive experiments have been conducted on the NSO dataset, and the results show that the detection and recognition ability of CS-YOLOv5 is superior to the comparison method mentioned in this paper. The experimental results also verify the effectiveness of the proposed improved methods in this paper for the problems needed to be solved, which can cope with the difficulties existing in space object detection, such as backlight environment, small scale, and huge difference of scale.

This paper has some main contributions, as follows:

We constructed the space dataset named NSO to solve the problem that data samples are difficult to obtain, including 10,574 samples annotated independently in COCO and YOLO format.

We propose CS-YOLOv5 for completing the space target detection task, which takes the YOLOv5 as the baseline to better detect small and weak objects in space.

We propose an augmentation strategy named contrast mosaic to enhance data complexity and diversity, which can make the network model applicable to more complex and difficult scenarios.

In CS-YOLOv5, we design the cross-layer context fusion module (CCFM) to extract the feature expression of multiple scales through parallel branches and integrate the context information to improve the detection performance of small objects.

In the CCFM, we design the adaptive weighting module (AWM) in combination with the attention mechanism to map the small-scale features sequentially to the features of the previous layer, which can further enhance the expression of features.

In CS-YOLOv5, we propose the spatial information enhancement module enable our model to not only enrich the multi-scale context information but also adaptively learn the weak spatial information of small objects that need to be protected, which can equalize the input features and extracts vertically and horizontally related features to small objects, capturing multi-directional spatial information.

We conduct extensive comparative experiments to verify that our method achieves higher detection accuracy on the NSO datasets. Ablation experiments confirmed that all parts of our method positively affect the improvement of the detection result.

2. Related Work

2.1. Data Augmentation

The essence of the data augmentation method is to expand the dataset and improve data quality by applying basic image processing operations such as flipping and adding noise to existing limited data so that the data can generate value equivalent to a larger amount of data. The SMOTE algorithm proposed by Chawla et al. [9] synthesizes new samples for small sample categories to solve the problem of sample imbalance. SMOTE maps the extracted image features to the feature space and selects a few adjacent samples after determining the sampling magnification. It randomly selects a connection line from them and randomly selects a point on the line as a new sample point, repeating this process until the sample is balanced. The Mixup proposed by Zhang et al. [10] performs basic data augmentation operations on two extracted images and concatenates the pixels to form a new sample in the form of averaging. Mixup can change the nonlinear relationship between each pixel of the data sample, blur the boundary of sample classification, and enhance the complexity of the training sample. The Cutmix proposed by Yun et al. [11] erases a portion of the pixel information on the image by covering it with a rectangular mask and then adds other sample information randomly in the erased region. The sample pairing proposed by Inoue et al. [12] usually randomly selects two pictures from the training set, performs basic data augmentation operations (such as random flipping, etc.), respectively, averages the pixels, and finally superimposes and synthesizes a new sample. It can significantly improve the classification accuracy of all test datasets. The copy enhancement method proposed by Kisantal et al. [13] copies and pastes small objects in the picture to increase the proportion of small objects in the dataset. To cope with the limited performance of the detector during training, the mosaic data augmentation method was proposed in YOLOv4 [14]—mosaic data augmentation splices together four pictures, each with its corresponding ground truth box. After splicing the four pictures, a new picture and its corresponding ground truth box are obtained. This new picture is then passed into the neural network for learning. It is equivalent to learning four pictures at a time. Due to objective reasons, the sample data collected in the space is not complete enough. To provide the model with sufficient data for learning and to improve the impact of small object size and low contrast on the detection accuracy, we propose contrasting a mosaic based on Mosaic. It performs histogram equalization, copy-paste, and flip operations on a sample image, respectively, and finally splices these four images to obtain a new image, effectively improving the diversity of datasets.

2.2. Multi-Scale Object Detection

In 2012, AlexNet, an important achievement of CNNs, successfully led people into the era of deep learning. As researchers continue to study deep learning technology, more and more object detection models appear in the public eye, mainly divided into two categories: the one-stage object detection model and the two-stage object detection model. Two-stage detectors such as R-CNN [15], Fast R-CNN [16], Faster R-CNN [17], and Mask R-CNN [18] have higher detection accuracy, but due to the large number of candidate areas, the detection speed is slow, making them unsuitable for practical application scenarios.

The You Only Look Once (YOLO) [19] series and single-shot multi-box detector (SSD) [20] represent typical single-stage object detection algorithms. Based on YOLOv1, Redmon et al. have continued to improve and propose the YOLOv2 [21] and YOLOv3 [22] detection algorithms. There are many normalization efforts in YOLOv2, including batch normalization (BN) technology and the introduction of an anchor frame mechanism. YOLOv3 employs darknet-53 as the backbone network, utilizes three different sizes of anchor boxes, and applies the Sigmoid function in the logical classifier to constrain the output between 0 and 1, enabling faster inference in YOLOv3. Rochkovskiy et al. added some practical techniques to the traditional YOLO and proposed the YOLOv4 algorithm, which replaces the ReLU activation function in the backbone network with the Mish activation function. Compared to the ReLU activation, the Mish function produces smoother images, achieving a better balance between detection speed and accuracy.

In the field of object detection, detecting small objects poses challenges in extracting effective feature information due to their limited pixel occupancy. Researchers have studied a series of methods to improve the performance of small object detection through network structure, training strategies, and data processing.

Tsung-Yi Lin et al. proposed the feature pyramid network (FPN) [23], which performs top-to-bottom upsampling on the features extracted from the bottom-up backbone network and merges the features extracted from the backbone network with the upsampled features to enhance the richness of feature details. the path aggregation network (PANet) proposed by Liu et al. [24] adds a bottom-up direction enhancement to FPN, allowing the top-level feature map to incorporate spatial information from the lower layers. Tan proposed the bidirectional feature pyramid network (BiFPN) [25] feature fusion network, combining top-down and bottom-up feature extraction processes into BiFPN layer. Additionally, in feature fusion, the weights are increased to control the proportion of different layers, and the input features are fully fused through multiple stacking BiFPN layer.

The attention mechanism module enhances the perception ability of features in both spatial and channel dimensions by weighting the useful information of the input features. The Squeeze-and-Excitation network (SENet) attention mechanism proposed by Jie et al. [26] squeezes and excites the channel information extracted from the input features to

obtain channel weights and then weights the input features. The Convolutional Block Attention Module (CBAM) attention mechanism module proposed by Sanghyun Woo et al. [27] weights the input features through the serial spatial attention module and the channel attention module, enhancing the utilization of spatial and channel information in the input features. The Coordinate Attention (CA) attention mechanism module was proposed by Qibin Hou et al. [28], which carries out a pooling operation to keep the position information of the input features, mines the attention information of the features, and activates the input features by obtaining mixed attention weights. Cheng et al. [29] proposed a rotationally invariant convolutional neural network called Rotation Invariant Convolutional Neural Networks (RICNN) to better detect small objects in remote sensing images. Liu et al. [30] proposed the Receptive Field Block Net (RFB-Net) combined with a multi-branch receptive field convolution module, using dilated convolution to further enhance the detection ability of small objects. Li et al. [31] constructed the enhanced feature pyramid network (eFPN) structure to reduce interference from the background on object detection in aerial images, improving detection accuracy. However, the context is not tightly connected, which can easily lead to information loss. Aiming at object aggregation in aerial images, Guo et al. [32] proposed the orientation perception feature fusion method to deal with the aggregation problem. Due to the complex network structure, more computing resources are needed. To improve object detection accuracy in UAV images, more detailed features need to be utilized, Wang et al. [33] used the inception lateral connection network (ILCN) structure based on feature pyramids to handle the scale changes of objects in aerial images. Tang et al. [34] proposed a remote sensing ship detection model N-YOLO, which, combined with a noise level classifier, achieves high-precision identification and incorporates an object potential area extraction module for more accurate positioning. Zhang et al. [35] proposed a reference-based method that uses the rich texture information of higher-solution reference images to compensate for the lack of detail in low-resolution images. Li et al. [36] achieved excellent performance in infrared target detection by extending and iterating the shallow CSP module of the feature extraction network and introducing multiple detection heads. Lu et al. [37] proposed an object detection method based on adaptive feature fusion and illumination-invariant feature extraction. Additionally, they introduced an adaptive crossscale feature fusion model to ensure the consistency of the constructed feature pyramid. Song et al. [38] proposed a multi-source deep learning object detection network based on the fusion of millimeter wave radar and vision by adding input channels and feature fusion channels to YOLOv5. They established two backbone networks for feature extraction, performed feature fusion at intermediate layers, and conducted detection as the final step.

Inspired by the mechanism of human perception, the attention mechanism is employed in object detection to focus and select useful information for the task. Although the existing research has achieved some achievements in the detection task of small objects, there are still problems of insufficient information extraction and fusion when dealing with weak and small space objects. In response to the above problems, we consider the existence of weak space objects in specific space scenes, so we combine the contextual information of objects and use the relationship between fragments and other objects or backgrounds to provide more effective information for detecting space objects. At the same time, we combine the attention mechanism to enable the network to focus on important information and reduce the interference of inaccurate information.

2.3. Space Object Detection

Due to the problems of small and dense objects, large scale differences, and low object contrast affected by stray light in space object images, detecting and tracking space objects using optical images remains a challenge in many space surveillance systems. Traditional space object detection methods perform well on single objects in simple backgrounds, but they may suffer from severe false alarms and missed detections in complex backgrounds, such as template matching, morphological operations, thresholding methods, and optical flow. Currently, space object detection methods are gradually moving toward deep learning.

Edward et al. [39] addressed the issue of limited feature extraction in traditional methods by proposing a star-galaxy classification framework that includes eight convolutional layers. This framework directly utilizes deep convolutional neural networks (ConvNets) on reduced and calibrated pixel values. Wu et al. [40] tackled the problem of low detection efficiency in traditional methods by proposing an artificial intelligence method for space object recognition called the two-stage convolutional neural network (T-SCNN). The T-SCNN consists of two stages: object locating and object recognition. Xiang et al. [41] introduced a fast spatial debris detection method based on grid learning. The image is divided into a grid of 14 imes 14 cells, and a grid-based fast neural network (FGBNN) is used to locate spatial debris within the grid. Wang et al. [42] improved the feature extraction structure based on the YOLOv3 network. Integrating shallow and deep features can enhance the network's detection capability for objects of different scales. Qualitative and quantitative experimental results demonstrate that the improved YOLOv3 network can accurately and effectively detect key components of space solar power plants. Jiang et al. [43] proposed a space object detection algorithm based on invariant star topological information, utilizing the relatively constant topological relationships between consecutive frames of celestial bodies. They quantized the invariant information into internal angle descriptors and designed a strategy to separate objects from celestial bodies. However, there is currently limited research on space object detection, and performance remains poor in complex scene detection tasks. Therefore, this study combines existing knowledge and object features to design a detection network that accurately detects space objects.

3. Proposed Method

3.1. Context Sensing-YOLOv5

Considering the simplicity and high efficiency, we adopt the YOLOv5 framework as a baseline. Yolov5 is mainly composed of input, backbone, neck, and head. This method first enhances the image at the input terminal and subsequently extracts feature maps of different scales from the backbone through structures such as focus and CSPDarknet. These feature maps are then fused in the neck, ensuring that each scale's feature map contains strong semantic and positional information. Finally, the feature maps are sent to the head for prediction. The CS-YOLOv5 we proposed mainly improves its neck and data enhancement parts; the overall framework is shown in Figure 3.

As shown in Figure 3, we add the CCFM between the backbone and the neck, which realizes the full fusion of different scopes of information through two parallel main branches. At the same time, in the CCFM, we design the AWM combined with the attention mechanism to efficiently fuse feature maps of various scales, thereby enhancing the information of each scale feature. Finally, we add the SIEM module between the neck and the head, which equalizes the input features and extracts vertically and horizontally related features to small objects, capturing multi-directional spatial environmental information.

3.2. Cross-Layer Context Fusion Module

To solve the problem that the model can perceive limited information when detecting small objects, we propose the CCFM, as shown in Figure 4. The CCFM obtains more semantic information from the feature map through two parallel branches and effectively mixes features of different scales to establish the relationship between the information of different scales.

As shown in Figure 4, the three-layer feature map (C4, C6, C9) extracted by Backbone from the input image will be enhanced by CCFM. The input CCFM features will obtain local and multi-scale semantic information through two different operations, respectively. The following formula can describe the approximate:

$$x1 = Conv_{3\times 3}(C),\tag{1}$$

$$x4 = Conv_{3\times3}(Conv_{3\times3}(Conv_{5\times5}(C))), \tag{2}$$

$$o1 = AWM [x1, o2],$$
 (3)

where *x*1 and *o*2 represent the feature map obtained by the first branch and the second branch, respectively; $Conv_{n \times n}(\cdot)$ represents the convolution operation; $n \times n$ represents the size of the convolution kernel; AWM [·] represents that the adaptive weighting module processes the feature map.



Figure 3. The framework of the proposed CS-YOLOv5.



Figure 4. The framework of the cross-layer context fusion module.

The branch responsible for extracting local information uses a 3×3 convolution to obtain favorable information about the small object. The second branch is designed concerning FPN, integrating three different scale features and reducing the imbalance among multi-scale features. This branch is divided into three smaller branches, and each branch models the features through a convolution layer. The convolution kernel sizes are five, three, and one from shallow to deep, with the step size being two. Through the AWM, this branch fuses the feature information of the output of each convolution kernel from smallest to largest. The AWM can nonlinearly fuse the representation information. Finally, the outputs of these two branches are fused by the AWM. The cross-layer context fusion module realizes the integration of context information of different scales and strengthens the nonlinear correlation between local information and multi-scale information to improve the detection performance of small space objects.

3.3. Adaptive Weighting Module

Aiming to address the issue of feature dilution caused by multiple upsampling during the feature fusion process and the impact of redundant information generated during the fusion process on the detection accuracy, we propose an adaptive weighting module, which can assist the network in effectively detecting dense and small objects. The structure of the AWM, as shown in Figure 5, combines the attention mechanism with sub-pixel convolution to achieve pixel-wise enhancement of feature maps.



Figure 5. Illustration of the adaptive weighting module.

During the feature fusion process, if the feature maps of different sizes are simply concatenated, the inaccurate spatial position information from the small-scale feature map will inevitably be introduced into the large-scale feature map, reducing the accuracy of spatial position information in the large-scale feature map and negatively impacting small object detection. To avoid the interference of high-level coarse-grained location information on the underlying fine-grained location information, we designed a weighting module focusing on semantic information, namely an adaptive weighting module. This module maps the information of small-scale features to the features of the previous scale through the attention mechanism, effectively utilizing the valuable information of each scale feature.

As shown in Figure 5, we input two adjacent feature maps from different layers in this module. We perform operations such as sub-pixel convolution on the small-scale coarse-grained feature map to obtain features that match the size of the large-scale fine-

grained feature maps. Sub-pixels are tiny pixels that exist between two actual pixels. Sub-pixel convolution can maximize the use of tiny pixels around a pixel on the image to achieve more refined interpolation calculation. Next, we apply adaptive max pooling and adaptive average pooling on this feature map in the W and H dimensions, respectively, to compress the feature map. Then, we pass the compressed feature maps through fully connected layers, the ReLU activation function, and another fully connected layer, followed by element-wise addition. This process automatically captures the interdependencies between the maximum and average features within the channels and spatial dimensions. Finally, the obtained feature is passed through the Sigmoid activation function to obtain the final channel attention. The channel attention is multiplied with the original feature map to assign different weights to each channel of the feature map, achieving a weighted fusion of features from different scales and obtaining the final output. The adaptive weighting module (AWM) can map the small-scale feature map to the feature map of the previous scale while avoiding feature dilution caused by multiple upsampling operations and enhancing the texture features of small objects.

3.4. Spatial Information Enhancement Module

To solve the problem that the spatial information of small objects is easily lost in the process of multiple convolutions, we proposed the SIEM to strengthen the learning of spatial information from both horizontal and vertical directions. The structure is shown in Figure 6.



Figure 6. Illustration of the spatial information enhancement module.

As shown in Figure 6, the SIEM contains two main branches and a skip connection, and each branch contains two branches and a skip connection. The feature maps of three scales (C20, C24, and C28) generated by the top-down pathway feature fusion are sent to the SIEM. The input feature map is equally divided on the channel to obtain feature maps Ci_1 and Ci_2:

$$Ci \in \mathbb{R}^{W \times H \times 2C} \to Ci_1 \in \mathbb{R}^{W \times H \times C}, \tag{4}$$

$$Ci \in \mathbb{R}^{W \times H \times 2C} \to Ci_2 \in \mathbb{R}^{W \times H \times C},$$
(5)

where *W* and *H* represent the scale of a feature map; *C* represents the channel number of a feature map; \rightarrow represents the operation of equalizing the feature map by channel.

To make full use of the characteristic information learned at each stage and enhance the interaction of information in different directions and channels, a series strategy is added between different branches based on parallel connection; that is, the output and original features of one branch are taken as the input of another branch at the same time. This way can ensure that the network can fully use the different feature information in the horizontal and vertical directions and increase the diversity of features.

Split and Fuse. Ci_1 and Ci_2 will pass through two branches, respectively. The first branch can map the spatial relative information in the horizontal direction. First, 1×1 convolution is used to reduce the number of channels. Then 1×3 convolution is used to perform one-dimensional row-dimensional convolution and enhance the relationship between feature points in horizontal space. The output of the first branch is spliced with Ci_1 or Ci_2 in the channel direction as the input of the second branch. The second branch can map the spatial relative information in the vertical direction. This branch is similar to the first branch, only replacing the 1×3 convolution kernel with the 3×1 convolution kernel to perform one-dimensional column-dimensional convolution and enhance the relationship between feature points in vertical space. Finally, the outputs of the two branches are spliced, and the information obtained by the branches is fused through a skip connection. We concatenate the outputs of the two main branches and then fuse all the information of the two directions of the object to transmit spatial information conducive to small object detection.

3.5. Contrast Mosaic Data Augment

Due to objective reasons, the collected sample data in space is incomplete. Meanwhile, we have noticed that the scale of debris in space is small, and the contrast is low due to the influence of the Earth and the sun as backgrounds. To provide the model with sufficient data for learning and to improve the impact of small object size and low contrast on the detection accuracy, we proposed a contrasting mosaic based on Mosaic. The process is shown in Figure 7.



Figure 7. The process of contrast mosaic data augment.

We randomly select an image from the dataset and, respectively, apply three different operations: histogram equalization, copy-paste, and turnover. Histogram equalization can enhance image contrast and improve the distinguishability of small objects in a backlight environment. We use copy-paste to segment, capture and paste small objects randomly to enhance the proportion of small and medium objects in the image. Finally, we choose how to turn over to transform the original image. By combining the images generated from these three operations with the original image using data augmentation techniques, we effectively improve the diversity of the dataset without altering the original object scales. Contrast mosaic mixes four training images so that the trained model can achieve the best detection effect for targets in different complex environments or scales, improving the detector's robustness. A few samples generated by the contrast mosaic are shown in Figure 8.



Figure 8. Some samples were generated by contrast mosaic data augmentation.

4. Results

4.1. Datasets

We evaluate our proposed method on the NSO dataset that we have simulated against a real spatial background and briefly introduce it as follows.

To solve the problem of hard obtaining the real space object dataset, we construct a space target simulation system, which can simulate the images obtained by observing space target objects from different continuously changing perspectives under different angles. The system mainly comprises a perspective motion modeling module, a scene rendering module, and an image post-processing module. The quality of the simulated images produced by this system is very high, which can provide a good data foundation for the research of space-based observation equipment detection, space-based target detection, identification, and analysis technology. We simulated the space environment of three types of objects: debris, satellites, and meteorites, by using the space object simulation system. To obtain data that better meets our requirements, we densely arranged the debris in space and adjusted the distance between the objects and observation points to change the object size. Then, we recorded the simulation process of the object's flight around the Earth from different observation angles. The farther the object is from the observation point, the smaller its size. To ensure high similarity between adjacent images, we extracted frames from recorded videos at intervals of 0.5 s. The resulting data sample contains 233,687 instances of 3 object categories in 10,574 images with a resolution of 2560×1440 . Some samples of the NSO dataset are shown in Figure 9. The red frame is a magnified display of the aggregated objects.

The number distribution of the three categories of objects is shown in Figure 10. The NSO dataset contains 84,742 small objects (area < 1024), 108,610 medium-scale objects (1024 < area < 9216) and 40,335 large objects (area > 9216). The average area ratio of the debris in the image is 0.45%, the satellite is 7.25%, and the meteorolite is 0.46%. The NSO dataset is divided into training, validation, and test sets according to a proportion of 8:1:1.



Figure 9. Some samples of the NSO dataset.



Detection category distribution

4.2. Implementation Details

Extensive experiments are implemented on Ubuntu16.04, the GPU adopts Nvidia GeForce GTX 1080Ti, and the CPU adopts Intel Core i7-11700K@3.80 GHz. The simulation experiment uses Pytorch 1.7.0, the development environment is Python 3.7, and the CUDA version is 10.2. In the course of training, we use pre-trained weights trained on the MS COCO dataset to initialize the parameters of the network. During the experiment this paper uses MMdetection [44] for the experiment. In the stage of training, stochastic gradient descent (SGD) is used as the optimizer, the batch size is 2, the original learning rate is 0.002,

Figure 10. The number distribution of NSO categories.

the momentum is 0.9, the weight decay is 0.0001, and each group of models is trained for 12 cycles.

4.3. Metrics

Two commonly used metrics, average precision (AP) and mean average precision (mAP), are used to evaluate the performance of CS-YOLOv5 in this paper. The relevant calculation formula is:

$$AP = \int_{0}^{1} P dR, \tag{6}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N},\tag{7}$$

where *P* and *R* represent precision and recall, respectively; *N* represents the number of detection categories. Average precision is related to precision and recall. *Precision* refers to the ratio between the number of positive samples correctly predicted in the prediction dataset and the number of positive samples predicted by the model. *Recall* rate refers to the ratio between the number of positive samples predicted correctly and the number of actual positive samples in the forecast dataset.

$$Precision = \frac{TP}{TP + FP'}$$
(8)

$$Recall = \frac{TP}{TP + FN'}$$
(9)

where *TP* is the number of positive samples predicted to be positive, which is the number of detection frames with the intersection of union (IoU) ≥ 0.5 . *FP* is the number of negative samples predicted as positive samples, which is the number of detection frames with IoU < 0.5, or the number of redundant detection boxes that detect the same object. *FN* is the number of predicted positive samples as negative samples. Mean Average Precision is the average of *AP* with IoU thresholds between 50% and 95% in COCO-style, and *AP*₅₀ (*AP* for IoU threshold 75%), *AP*_S (*AP* for small objects), *AP*_M (*AP* for medium objects) and *AP*_L (*AP* for large objects) are used to illustrate the results.

4.4. Ablation Experiments

In this section, we will conduct extensive experiments to verify the gain effect of the data augmentation method and feature fusion modules on the model proposed in this paper. Additionally, we will discuss the results in detail as follows.

To verify the improvement of models by the contrast mosaic data augmentation method, extensive experiments are conducted on seven state-of-the-art methods and our method for comparison, including YOLOv5, ATSS [45], FSAF [46], FCOS [47], TOOD [48], RetinaNet [49], VFNet [50] and CS-YOLOv5. The experimental result is shown in Table 1; the better results are marked in bold.

It can be seen from the results that all methods have different degrees of map enhancement after using data augmentation. In addition, the AP_S of these methods is significantly improved, leading to a gain from 5.0% to 12.3%, which indicates that contrast mosaic can achieve a balanced improvement for the network to detect space objects. For example, we selected YOLOv5 as the baseline in this paper; the contrast mosaic data augmentation brings a 3.4% gain to its mAP. Compared with other networks, this gain increases at a moderate level. In addition, the AP_S of the YOLOv5 achieves the best result of all methods. At the same time, the AP_M and AP_L also obtain the best results among all methods. In addition to the existing comparison model, we use the proposed model to verify the validity of the contrast mosaic data augmentation method. As shown in Table 1, the mAP of our method reaches 69.6%, and the AP_{50} , AP_{75} , AP_S , AP_M , and AP_L all have improvement. Experimental results show that the method proposed in this paper can increase the number and diversity of training samples and improve the generalization ability and robustness of the model through three different processing of the original dataset and then splicing the three generated images with the original images.

Methods	Backbone	Contract Mosaic	1	AP (%), IoU			AP (%), Area		
		Contrast Mosaic	0.5:0.95	0.5	0.75	S	М	L	
YOLOv5	CSPDarknet-53	-	62.9	87.0	74.4	36.2	51.3	72.9	
		\checkmark	66.3	89.5	75.6	47.4	76.8	87.7	
	ResNet-50	-	61.1	86.2	72.0	30.3	59.2	76.3	
AISS		\checkmark	64.3	88.7	76.5	42.6	73.6	85.5	
FSAF	ResNet-50	-	51.8	78.4	63.5	30.1	51.7	63.6	
		\checkmark	60.6	84.5	69.0	40.3	68.8	78.4	
FCOS	ResNet-50	-	57.1	84.8	65.8	26.7	56.4	74.3	
		\checkmark	63.1	90.9	73.7	33.5	71.1	78.8	
TOOD	ResNet-50	-	64.3	88.6	74.6	36.2	55.6	75.7	
		\checkmark	65.8	90.4	77.2	46.8	76.1	87.5	
RetinaNet		-	51.9	77.1	67.6	30.1	55.0	63.5	
	KesiNet-50	\checkmark	59.1	84.9	71.3	35.1	69.2	78.7	
VFNet	D N (50	-	56.9	84.9	64.3	32.6	55.9	64.9	
	KesiNet-50	\checkmark	62.4	87.5	73.8	40.4	65.8	71.2	
		-	67.8	91.6	79.4	48.8	68.1	79.5	
CS-YOLOv5	CSPDarknet-53	\checkmark	69.6	93.8	80.7	56.3	82.5	89.6	

Table 1. Ablation experiment of the contrast mosaic on the NSO dataset.

We conduct extensive experiments on the NSO dataset to verify the effectiveness of each improvement in CS-YOLOv5. We choose YOLOv5 as the baseline and added the CCFM, AWM, and SIEM to compare the performance changes. When using CCFM without AWM, we replace the AWM with simple interpolated upsampling. The experimental result is shown in Table 2; the best results are marked in bold. At the same time, *FPS* is used as the evaluation standard of model speed.

Methods	CCFM	AWM	SIEM	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	FPS
YOLOv5	-	-	-	62.9	87.0	76.4	36.2	51.3	72.9	58.8
Ours	\checkmark	-	-	64.7	89.4	78.5	38.4	60.2	76.0	56.3
Ours	\checkmark	\checkmark	-	65.5	89.9	79.6	41.0	65.4	78.1	54.9
Ours	-	-	\checkmark	65.3	88.8	78.2	42.2	65.6	76.7	55.4
Ours	\checkmark	\checkmark	\checkmark	67.8	91.6	79.4	48.8	68.1	79.5	48.4

Table 2. Ablation experiment of the components on the NSO dataset.

According to the experimental results shown in Table 2, the CCFM improves the mAP of the baseline by 1.8% and the AP_S by 2.2%. In addition, other metrics also gained varying degrees of improvement. This confirms that the CCFM effectively integrates multi-scale information of the same feature layer conducive to detecting small objects; this improvement is reasonable. After adding the AWM to the baseline and CCFM, the *mAP* and the AP_S are significantly improved relative to the baseline, leading to a gain of 2.6% and 6.0%. At the same time, the AP_M and AP_L also increased by 14.1% and 5.2%, achieving relatively large gains. These results show that using this module to transfer semantic information between feature maps of different sizes by computing attention can effectively improve the detection performance of small objects, and improve the detection performance of large and medium-scale objects, to improve the overall performance. It proves the indispensability of the AWM. When only the SIEM is added to the baseline model, the network improved *mAP* less significantly than the CCFM and the AWM simultaneously, but it performed better in AP_S than the other two modules. For mAP and AP_S , the SIEM improves the baseline by 2.4% and 6.0%, respectively. The SIEM can enrich the texture information of space objects and avoid excessive information loss in the process of feature

transfer, thereby improving the detection ability of small objects. When all three modules are inserted into the baseline, all indicators achieve the best results. The *mAP* and *AP*_S increased by 4.9% and 12.6%, respectively. When the three modules work together, the effect is better than any module working alone. After experimental verification, the corresponding improvement measures proposed in this paper for the difficulties in space weak and small object detection tasks can improve the detection performance of the baseline model in all aspects. Regarding detection speed, YOLOv5 has an *FPS* of 58.8, while CS-YOLOv5 has a lower *FPS* than YOLOv5, but it still has real-time detection capability. In summary, the method proposed in this paper can achieve accurate and fast detection of weak space and small objects. Meanwhile, we calculated the indirect cost required to improve the model, and the results are shown in Table 3. Params represent the parameter quantity of the model, which can measure the model's size. GFLOPS is the amount of calculation that can measure the complexity of the model. CS-YOLOv5 increases the parameter count by 9272061M and the calculations by 11.1GFLOPS compared to the baseline model. Although the proposed model adds some computational cost, the performance is effectively improved.

Table 3. Result of indirect cost statistics.

Methods	FPS	Params	GFLOPS
YOLOv5	58.8	7,056,607	16.3
CS-YOLOv5	48.4	16,328,668	27.4

4.5. Performance Results

We compare our proposed CS-YOLOv5 with nine state-of-the-art methods, such as YOLOv5, ATSS, FSAF, FCOS, TOOD, RetinaNet, VFNet, GFL [51], and PAA [52]. The comparison model and CS-YOLOv5 were trained under the same environment and dataset. To reflect the effectiveness of the proposed methods and reduce the impact on detection performance due to the difference in the backbone, the comparison methods involved in this paper, except CS-YOLOv5 and YOLOv5, use ResNet-50 as the backbone, which is more objective. The experimental result is shown in Table 4; the best results are marked in bold.

Table 4. Performance evaluation of different methods on the NSO dataset.

Method	Backbone	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv5	CSPDarknet-53	62.9	87.0	74.4	36.2	51.3	72.9
ATSS	ResNet-50	61.1	86.2	72.0	30.3	59.2	76.3
FSAF	ResNet-50	51.8	78.4	63.5	30.1	51.7	63.6
FCOS	ResNet-50	57.1	84.8	65.8	26.7	56.4	74.3
TOOD	ResNet-50	64.3	88.6	74.6	36.2	55.6	75.7
RetinaNet	ResNet-50	51.9	77.1	67.6	30.1	55.0	63.5
VFNet	ResNet-50	56.9	84.9	64.3	32.6	55.9	64.9
GFL	ResNet-50	63.7	88.4	72.5	36.0	54.4	76.0
PAA	ResNet-50	59.1	84.9	66.3	33.1	51.2	78.7
CS-YOLOv5	CSPDarknet-53	67.8	91.6	79.4	48.8	68.1	79.5

It can be seen from Table 4 that the *mAP* of the CS-YOLOv5 on the NSO dataset is 67.8%, which is 4.9% higher than YOLOv5. Compared with the above competing methods, the results of metrics of our method have significant advantages, especially the AP_S shows a 12.6 percent improvement compared with the baseline model. In addition, the AP_M and AP_L of CS-YOLOv5 increased by 16.8% and 5.6%, respectively. Therefore, the proposed method can handle complex spatial environments and multi-scale objects, achieving good detection performance. Some examples of the detection results on the NSO dataset are shown in Figure 11.



Figure 11. Sample testing results of CS-YOLOv5.

5. Discussion

This paper proposed several improvements based on YOLO to detect weak and small objects in space. The proposed CS-YOLOv5 can achieve high-precision detection of space-weak and small objects, low-contrast objects, clustered objects, and multi-scale objects. Compared to several state-of-the-art methods, the results of this research have been improved. The research is currently stuck in the simulation and verification stage, but in space, the models are mostly deployed on onboard devices with limited resources and computing power. Therefore, the models need to have small size and fast speed characteristics while achieving a certain degree of accuracy. Hence, the next research stage will focus on building a lightweight model to make the detection process fast and accurate.

6. Conclusions

In this paper, we propose a context sensing-YOLOv5 optimized for small and weak debris in space. The proposed method is applied to the YOLOv5 to solve the problems of small and weak object detection in space. First, a cross-layer context fusion module is proposed to realize the integration of context information of different scales and strengthens the nonlinear correlation between local information and multi-scale information. Secondly, an adaptive weighting module based on the attention module is proposed to transfer information from a small to a large scale. This can avoid the feature dilution problem caused by multiple upsampling and enhance the texture features of small objects. Thirdly, a spatial information enhancement module is proposed to take different convolution measures for different channels to obtain detailed information in different directions. We propose an augmentation strategy named contrast mosaic to enhance data complexity and diversity, which can make the network model applicable to more complex and difficult scenarios. We conduct ablation experiments on the NSO dataset to verify the effectiveness of each improvement in CS-YOLOv5. Our experiment results show that CS-YOLOv5 performs better than the other object detection methods compared in this paper and can satisfy the requirements of spatial objection detection in different situations, such as dense distribution, extremely small scale, and large-scale difference.

Author Contributions: All authors contributed to this manuscript. Methodology, Experimental results analysis, writing original draft, Y.Y.; Writing—review and editing, H.B. and P.W.; Data curation, H.G. and T.D.; Draw the diagram, W.Q.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC), grant number U2031138 and the National Defense Science and Technology 173 Program Technology Field Fund of China (2022-JCJQ-JJ-0395).

Acknowledgments: The authors would like to thank the anonymous reviewers for the constructive suggestions.

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

References

- 1. LEGEND: 3D/OD Evolutionary Model. Available online: https://orbitaldebris.jsc.nasa.gov/modeling/legend.html (accessed on 14 September 2022).
- Yang, Y.; Lin, H. Automatic Detecting and Tracking Space Debris Objects Using Active Contours from Astronomical Images. *Geomat. Inf. Sci. Wuhan Univ.* 2010, 35, 209–214.
- 3. Lowe, D. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 2004, 60, 91–110. [CrossRef]
- 4. Richard, O.D.; Peter, E. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM* 1972, 15, 11–15.
- 5. Freund, Y.; Schapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
- 6. Felzenszwalb, P.; Girshick, R.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [CrossRef] [PubMed]
- 7. Ultralytics. YOLOv5. Available online: https://github.com/ultralytics/yolov5 (accessed on 1 November 2021).
- 8. Park, T.; Märtens, M.; Lecuyer, G.; Izzo, D. SPEED+: Next-generation dataset for spacecraft pose estimation across domain gap. In Proceedings of the IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 5–12 March 2022; pp. 1–15.
- 9. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W. SMOTE: Synthetic minority over-sampling technique. *Artif. Intell. Res.* 2022, 16, 321–357. [CrossRef]
- 10. Zhang, H.; Cisse, M.; Dauphin, Y.; Lopez-Paz, D. Mixup: Beyond empirical risk minimization. In Proceedings of the ICLR 2018 Conference Blind Submission, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–13.
- Yun, S.; Han, D.; Oh, S.; Chun, S.; Choe, J.; Yoo, Y. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6022–6031.
- 12. Inoue, H. Data Augmentation by Pairing Samples for Images Classification. arXiv 2018, arXiv:1801.02929.
- 13. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. *arXiv* 2019, arXiv:1902.07296.
- 14. Bochkovskiy, A.; Wang, C.; Liao, H. Yolov4: Optimal Speed and Accuracy of Object Detection. arXiv 2020, arXiv:2004.10934.
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.
- 16. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13–16 December 2015; pp. 1440–1448.
- 17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* 2015, arXiv:1506.01497. [CrossRef] [PubMed]
- 18. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
- 19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
- 21. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
- 22. Redmon, J.; Farhadi, A. Yolov3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767.
- 23. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
- 24. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
- Tan, M.; Pang, R.; Le, Q. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.

- Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
- 27. Woo, S.; Park, J.; Lee, J.; Kweon, I. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland; Munich, Germany, 2018; pp. 3–19.
- 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 (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.
- 29. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 7405–7415. [CrossRef]
- Liu, S.; Huang, D.; Wang, Y. Receptive field block net for accurate and fast object detection. In Proceedings of the European conference on computer vision, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland; Munich, Germany, 2018; pp. 385–400.
- Li, C.; Xu, C.; Cui, Z.; Wan, D.; Jie, Z.; Zhang, T.; Yang, J. Learning object-wise semantic representation for detection in remote sensing imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 20–27.
- 32. Guo, Y.; Xu, Y.; Li, S. Dense construction vehicle detection based on orientationaware feature fusionconvolutional neural network. *Autom. Constr.* **2020**, *112*, 103124. [CrossRef]
- Wang, J.; Ding, J.; Guo, H.; Cheng, W.; Pan, T.; Yang, W. Mask OBB: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote Sens.* 2019, *11*, 2930. [CrossRef]
- Tang, G.; Zhuge, Y.; Claramunt, C.; Men, S. N-YOLO: A SAR Ship Detection Using Noise-Classifying and Complete-Target Extraction. *Remote Sens.* 2021, 13, 871. [CrossRef]
- 35. Zhang, Z.; Wang, Z.; Lin, Z.; Qi, H. Image super-resolution by neural texture transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7982–7991.
- Li, S.; Li, Y.; Li, Y.; Li, M.; Xu, X. YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. *IEEE Access* 2021, 9, 141861–141875. [CrossRef]
- Lu, Y.; Gao, J.; Yu, Q.; Li, Y.; Lv, Y.; Qiao, H. A Cross-Scale and Illumination Invariance-Based Model for Robust Object Detection in Traffic Surveillance Scenarios. *IEEE Trans. Intell. Transp. Syst.* 2023, 22, 1–11. [CrossRef]
- Song, Y.; Xie, Z.; Wang, X.; Zou, Y. MS-YOLO: Object Detection Based on YOLOv5 Optimized Fusion Millimeter-Wave Radar and Machine Vision. *IEEE Sens. J.* 2022, 22, 15435–15447. [CrossRef]
- Kim, E.J.; Brunner, R.J. Star–galaxy classification using deep convolutional neural networks. Mon. Not. R. Astron. Soc. 2016, 464, 4463–4475. [CrossRef]
- Wu, T.; Yang, X.; Song, B.; Wang, N.; Gao, X.; Kuang, L.; Nan, X.; Chen, Y.; Yang, D. T-SCNN: A Two-Stage Convolutional Neural Network for Space Target Recognition. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1334–1337.
- Xiang, Y.; Xi, J.; Cong, M.; Yang, Y.; Ren, C.; Han, L. Space debris detection with fast grid-based learning. In Proceedings of the 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI), Chongqing, China, 28–30 November 2020; pp. 205–209.
- 42. Wang, G.; Lei, N.; Liu, H. Improved-YOLOv3 network for object detection in simulated Space Solar Power Systems images. In Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 7177–7181.
- Jiang, F.; Yuan, J.; Qi, Y.; Liu, Z.; Cai, L. Space target detection based on the invariance of inter-satellite topology. In Proceedings of the 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 17–19 June 2022; pp. 2151–2155.
- 44. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Lin, D. MMDetection: Open mmlab Detection Toolbox and Benchmark. *arXiv* 2019, arXiv:1906.07155.
- Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9759–9768.
- 46. Zhu, C.; He, Y.; Savvides, M. Feature selective anchor-free module for single-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 840–849.
- Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9627–9636.
- Feng, C.; Zhong, Y.; Gao, Y. TOOD: Task-aligned One-stage Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, NA, Canada, 10–17 October 2021; pp. 3490–3499.
- Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
- 50. Zhang, H.; Wang, Y.; Dayoub, F.; Sünderhauf, N. VarifocalNet: An IoU-aware Dense Object Detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 8514–8523.

- Li, X.; Wang, W.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 11632–11641.
- 52. Kim, K.; Lee, H. Probabilistic anchor assignment with iou prediction for object detection. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 355–371.

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.