electronics

Review

A Review: Remote Sensing Image Object Detection Algorithm Based on Deep Learning

Chenshuai Bai, Xiaofeng Bai and Kaijun Wu *

Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 11200726@stu.lzjtu.edu.cn (C.B.); 12221908@stu.lzjtu.edu.cn (X.B.)
* Correspondence: wkj@mail.lzjtu.cn; Tel.: +86-13099220306

Abstract: Target detection in optical remote sensing images using deep-learning technologies has a wide range of applications in urban building detection, road extraction, crop monitoring, and forest fire monitoring, which provides strong support for environmental monitoring, urban planning, and agricultural management. This paper reviews the research progress of the YOLO series, SSD series, candidate region series, and Transformer algorithm. It summarizes the object detection algorithms based on standard improvement methods such as supervision, attention mechanism, and multi-scale. The performance of different algorithms is also compared and analyzed with the common remote sensing image data sets. Finally, future research challenges, improvement directions, and issues of concern are prospected, which provides valuable ideas for subsequent related research.

Keywords: deep learning; optical remote sensing image; object detection; comparative analysis of performance

check for updates

Citation: Bai, C.; Bai, X.; Wu, K. A Review: Remote Sensing Image Object Detection Algorithm Based on Deep Learning. *Electronics* **2023**, *12*, 4902. https://doi.org/10.3390/electronics12244902

Academic Editors: Qiang Wang, Weihong Ren and Huijie Fan

Received: 2 November 2023
Revised: 17 November 2023
Accepted: 29 November 2023
Published: 6 December 2023

1. Introduction

Optical remote sensing image object detection has always been one of the research hotspots in the field of computer vision, and its main task is to locate and classify the objects of interest in remote sensing images [1]. As illustrated in Figure 1, the trained deep-learning algorithm is utilized to identify and predict the object of remote sensing images. The task of remote sensing image object detection is highly challenging and has broader application prospects, such as automatic driving [2], face recognition [3], pedestrian detection [4], medical detection [5], and so on. At the same time, object detection can also be used for image segmentation [6], image description [7], object tracking [8], action recognition [9], and other more complex computer vision tasks. Preprocessing, feature extraction, and object classification are typically the three processes of traditional remote sensing picture object detection techniques. First, the entire image is scanned using the sliding window approach, which may be configured with various aspect ratios to improve the capacity to recognize objects of various forms. This allows for the division of the image into multiple candidate regions.

Then, each sliding window is subjected to feature extraction. The matching technique scale-invariant feature transform [10] (SIFT), the [11] histogram of oriented gradient (HOG), and Haar features are examples of traditional features. These characteristics mainly depend on the image's texture, color, and scale. By extracting features, the object can be more accurately characterized. Finally, the sliding window is classified using conventional machine-learning classifiers like support vector machines (SVMs) and Adaboost to perform object detection. Using the retrieved features, the classifier will decide if the object item is present in the sliding window. Accurate object recognition can be achieved by training the classifier and learning the feature distribution of the object. In 2001, the birth of V-J detector was mainly used for face detection. The HOG+SVM approach first surfaced in 2006 and was primarily used to detect pedestrians. In 2008, Felzenszwalb presented the deformable part model (DPM) methodology [12]. For a long time before the deep-learning method matured, the DPM algorithm has been playing a role in the field of object detection. These are conventional image processing and computer vision-based item detection techniques. Traditional object detection algorithms for optical remote sensing images mainly rely on hand-designed feature extraction and traditional machine-learning classifiers. These methods have a strong interpretability of features such as texture, color, and scale of the object. However, traditional methods have some limitations in dealing with complex scenes, object scale changes, occlusion, and so on. Traditional approaches also have certain restrictions regarding computational correctness and efficiency due to the high resolution and broad coverage area of remote sensing images.

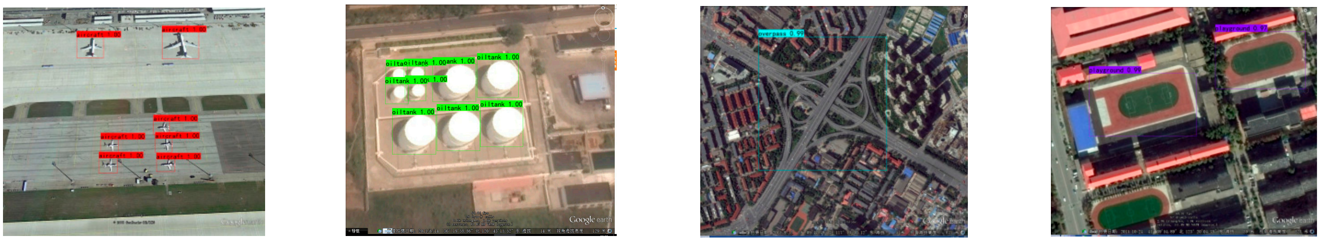

Figure 1. Object detection example diagram (The purpose of this figure is to show that we use the existing YOLOv8 algorithm model to perform object prediction and recognition in different categories of remote sensing image sample images of RSOD).

Deep-learning-based object recognition techniques have steadily emerged in recent years as computer vision and deep learning have advanced, which can better solve the limitations of traditional methods and have made significant breakthroughs in remote sensing image object detection.

When compared to the conventional method based on manual features, the object detection method based on deep learning has significant advantages. Traditional methods usually require manual designs of feature extractors, and the effect is limited under complex scenes and highly variable conditions. However, object detection methods based on deep learning can learn feature representations directly from the original image via end-to-end training without relying on hand-designed feature extractors. The Convolutional Neural Network (CNN) deep-learning model can extract rich feature information from images and capture semantic information at different levels, thus improving the accuracy and generalization ability of object detection.

Deep-learning-based object identification algorithms like Region-based Convolutional Neural Networks (Faster RCNN) [13], You Only Look Once (YOLO) [14], Single Shot MultiBox Detector (SSD) [15], and others have recently been popular research topics. These algorithms effectively address the balance between accuracy and efficiency in object detection by introducing different network structures and optimization methods. For example, Faster RCNN [13] used the Region Proposal Network (RPN) to generate candidate object frames before classifying and locating objects utilizing a series of convolutional and fully connected layers. However, YOLO [14] turned the object identification challenge into a regression issue and used a single CNN model to directly forecast the category and location of the object.

This paper provides a detailed summary of the research progress based on the YOLO series, SSD series, candidate region series, and Transformer algorithm. This study also explores the practical applications of these algorithms in the domain of optical remote sensing image object detection. This study involves a comparative investigation of various algorithms' performance, and the commonly used remote sensing image data sets are used. Additionally, the influence of common enhancement techniques, including supervised learning, attention mechanism, and multi-scale on the object detection algorithm is summarized. By conducting these comparisons and analyses, we can enhance our comprehension of the benefits and constraints associated with different algorithms. This will enable us to offer valuable insights and recommendations for future research endeavors, thereby fostering the advancement and utilization of this particular field.

2. Object Detection Algorithms for Optical Remote Sensing Images

2.1. Remote Sensing Object Detection Algorithm Based on Anchor Frame

2.1.1. YOLO Series Object Detection Algorithms

Due to their quick and precise capabilities, YOLO series object recognition algorithms have drawn a lot of attention. The first YOLO algorithm to be proposed was YOLOv1 [14]. Using an end-to-end single-stage network, it classified and located objects simultaneously, greatly increasing detection speed while retaining detection accuracy. YOLOv2 [16] improved YOLOv1 by introducing Darknet-19 as the backbone network and adopting the Anchor mechanism and multi-scale training strategy while using BN (Batch Normalization) to accelerate convergence and avoid overfitting. It had better recall and positioning accuracy and could adapt to multi-size image inputs. Based on YOLOv2, YOLOv3 [17] employed a more powerful darknet-53 backbone network and introduced the FPN structure to achieve multi-scale detection. It improved detection accuracy and speed by removing the pooling and fully connected layers and switching to logistic regression from softmax for classification. In addition, YOLOv3 provided a flexible version of tiny-darknet. YOLOv4 [18] introduced CSPDarknet53 as the backbone network with mosaic data enhancement and SPP+PAN structure to improve feature representation and cross-scale information fusion capabilities. It used GIOU_Loss instead of the Smooth L1 Loss function to improve the detection accuracy. The YOLOv5 [19] adopted CSPDarknet53 and Focus network based on YOLOv4 with Mosaic data enhancement. It used FPN (Feature Pyramid Network) and PAN (Pyramid Attention Network) structures for feature fusion and improved loss function and prediction frame filtering methods. Its lightweight model and high-accuracy object detection made it an excellent choice. The YOLOX [20] algorithm further improved its accuracy and speed via clever integration schemes such as data augmentation, anchor-free detection, and label classification. YOLOv6 [21] used EfficientRep Backbone and Rep-PAN Neck structure with an anchor-free paradigm and SIoU (Scale-Invariant Intersection over Union) bounding box regression loss optimization model. At the same time, it also utilized post-training quantization and quantitative perception training to improve the reasoning speed. YOLOv7 [22] was an improved object detection algorithm based on YOLOv5 and YOLOR, which introduced ELAN (Efficient and Lightweight Aggregation Network, ELAN), Rep (Reparameterization Convolution, Rep), and Auxiliary Head detection and other new techniques, and used YOLOX's SimOTA (Sim Optimal Transport Algorithm, SimOTA) strategy and YOLOV5's cross-grid search for label assignment. YOLOv7 was a worthwhile model to learn. YOLOv8 [23] was an improved version of YOLOv5 based on YOLOv5, which was fast, accurate, and easy to use for various object detection and tracking, instance segmentation, image classification, and pose estimation tasks. It adopted the overall structure of the CSPDarknet backbone network, the PAN-FPN neck, and the decoupled head. It improved the details such as using Context Cascaded Fully Convolutional Network C2f (Context Cascaded Fully Convolutional Network, C2f) module instead of the C3 (Context Cascaded Convolutional Network, C3) module and removing some convolutional structures in PAN-FPN. Compared with YOLOv5, YOLOv8 was fundamentally different, which was a model based on the idea of anchor-free. The object detection algorithm has undergone several iterations and upgrades of YOLOv1, YOLOv2, YOLOv3, YOLOv4, YOLOv5, YOLOX, YOLOv6, YOLOv7, and YOLOv8, which mainly includes the use of different backbone networks, the introduction of the Anchor mechanism, multi-scale training strategy, BN, logistic regression instead of softmax, FPN structure, CSP (CSP, Cross Stage Partial) Darknet, Mosaic data enhancement, SPP (Spatial Pyramid Pooling, SPP) and PAN structure, GIoU_Loss (Generalized Intersection over Union Loss, GIoU_Loss), Focus network, anchor-free detection, and other new technologies and methods, which significantly improve the speed and accuracy of object detection, and gradually move towards lightweight and efficient. Among them, YOLOv5 and YOLOX are relatively novel and excellent models, while YOLOv8 is an improved model based on an anchor-free idea. Table 1 provides a summary of the features, benefits, and drawbacks of YOLOv1-v8:

Table 1. Comparison of YOLOv1-v8 versions.

Version	Time	Characteristics	Weakness	Advantage
YOLOv1 [14]	2015	Fully convolutional networks, high real-time, simple, and effective.	Poor positioning accuracy, and poor detection of small objects.	Fast, easy to implement and deploy.
YOLOv2 [16]	2016	Anchor box, multi-scale prediction, and Darknet-19 are used as the basic network with significant accuracy improvement.	Detection of small objects remains difficult.	Fast speed, high overall accuracy, and good robustness.
YOLOv3 [17]	2018	Multi-scale prediction, FPN, better objecting	Relatively slow, higher computing resources are needed.	High detection accuracy, the effect of small object detection is improved, and strong robustness.
YOLOv4 [18]	2020	Powerful detection performance, faster speeds, higher accuracy, CSPDarknet53 network.	Requires higher computational resources and higher complexity.	Excellent detection accuracy, good detection effect on small objects and occluded objects, and strong robustness.
YOLOv5 [19]	2020	Lightweight network, fast speed, high precision, easy to train and deploy.	Relatively new, there may be some instability and room for improvement.	High speed and high accuracy, efficient performance and training deployment efficiency, multi-language support, and smaller model volume.
YOLOX [20]	2021	Decoupled head, anchor-free, and SimOTA.	Currently, there are only 640 × 640 pre-trained weights.	Better performance at lower image resolutions.
YOLOv6 [21]	2022	Support model training, reasoning, and multi-platform deployment, improvement, and optimization of network structure and training strategy.	High false detection rate, version maintenance update speed is too slow, not suitable for industrial fields.	Full-chain industrial application requirements, improved and optimized network structure, and algorithm level.
YOLOv7 [22]	2023	Better precision and rate, with high accuracy and fast detection speed.	Newer knowledge may be difficult for some users to learn.	It has better accuracy and speed while ensuring accuracy and can process video and images in real time.
YOLOv8 [23]	2023	Fast and efficient, high accuracy, supports multi-category object detection, suitable for real-time applications.	Poor detection of small objects, A large amount of training data, and training time are required.	Fast speed, high accuracy, fitting real-time scenarios, supporting multi-category object detection.

2.1.2. Remote Sensing Object Detection Algorithm of the YOLO Series

The architecture of the optical remote sensing object recognition algorithm based on regression analysis is shown in Figure 2. Due to its benefits of lightweight deployment, the YOLO series of object identification algorithms has the potential for a wide range of applications in the field of remote sensing. To increase the functionality of YOLO in the area of remote sensing photos, researchers have experimented with a variety of improvement and optimization techniques. For example, Tang et al. [24] proposed a moving object detection method based on dual-beam SAR that utilized two independent YOLO networks to fuse the sub-image results and improve detection accuracy. In tasks like aircraft detection in remote sensing photos [25], multi-source underwater object wake [26], and remote sensing building detection, the technique based on YOLOv5 proved accurate and effective [27]. FRN (Feature Refinement Network, FRN)-YOLO [28] was a feature refusion network for remote sensing object detection with high accuracy and robustness. In addition, there are some studies devoted to improving the performance of small object detection. Wei et al. [29] improved YOLOX and introduced a bilateral attention mechanism to effectively detect small-sized objects and reduce the false detection rate. A lightweight YOLOv4 algorithm with excellent efficiency and accuracy in remote sensing image recognition was proposed by Ma et al. [30]. It extracted feature information using the residual module and FPN and further improved performance via data enhancement and transfer learning. In addition, some researchers have considered the characteristics and application requirements of remote-sensing images. Li et al.'s [31] accurate and effective identification was made possible by tailoring the network structure and loss function to the properties of remote sensing images. The performance was then further enhanced using methods such as data augmentation and a priori box optimization. In order to improve performance, Hong et al. [32] accomplished multi-scale ship detection in synthetic aperture radar (SAR) and optical pictures and introduced multi-scale feature fusion. Yang et al. [33] improved the accuracy and robustness by redesigning the network structure and adjusting the loss function. In addition to the above methods, there are many improvements and applications based on the YOLO algorithm [34–42]. The related ideas of the above algorithms are shown in Table 2. Although the YOLO series algorithms have achieved significant advancements in the field of remote sensing image object detection, they encounter challenges due to the limited accuracy in detecting small objects, distant objects, and objects with typical contrast. Furthermore, these algorithms struggle to adapt to variations in data across different areas, seasons, and weather conditions. This paper discusses the various challenges that arise in ensuring the robustness and generalization capabilities of algorithms. In addition, in some application scenarios, it may be imperative to further improve and optimize the algorithm to improve the detection accuracy and reduce the false detection rate. Therefore, further research is needed to solve these problems and improve the performance of the algorithm.

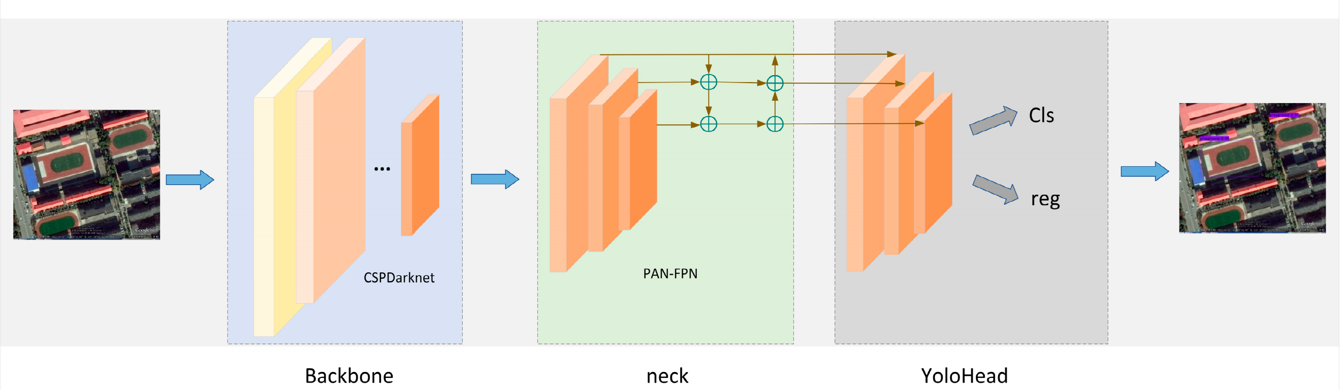

Figure 2. Basic process of optical remote sensing object detection based on regression analysis (the architecture diagram includes input and output; backbone: CSPDarknet backbone feature extraction network; neck: Path Aggregation Network (PAN)-Feature Pyramid Network (FPN) feature enhancement module and YOLOHead detection head).

Table 2. Comparison of object detection algorithms for remote sensing images based on YOLO series.

Literature	Paper Highlights	Applicability
Tang et al. [24]	Moving object detection method based on the YOLO, processing of dual-beam SAR images	Applying for moving object detection in dual-beam SAR images.
Jindal et al. [25]	Using the YOLOV5 architecture to realize aircraft detection in remote sensing images.	Applying for aircraft inspection tasks.
Shi et al. [26]	The improved YOLOv5 realizes the wake detection of underwater objects based on multi-source images.	Applying for the field of underwater object wake detection and having certain practical value.
Ding et al. [27]	Remote sensing image building detection with high accuracy and robustness.	Applying for remote sensing image building detection and using in urban planning, resource management, and other fields.
Sun et al. [28]	Feature re-fusion extracts details, reducing false detections.	Fast and accurate detection of the object, providing important data support.
Wei et al. [29]	Utilizing bilateral attention, detailed information about small objects in remotely sensed images is effectively captured.	The network focuses on improving small object detection.
Ma et al. [30]	Optimizing the network structure and parameters and achieving a lightweight.	Providing accurate object detection results.
Li et al. [31]	Data enhancement techniques are employed.	Real-time monitoring of ships, improving collection efficiency and safety.
Hong et al. [32]	Multi-scale detection, improved network structure and loss function, etc.	Maritime traffic monitoring, maritime patrol and island defense, marine resource management, and other areas.
Yang et al. [33]	The GIoU evaluation metric is introduced in the loss function.	Applying for aviation monitoring, military reconnaissance, disaster monitoring, and other fields.
Xin et al. [34]	Optimizing the characteristics of remote sensing images, such as high resolution and complex backgrounds.	Large-scale and complex remote sensing image data, with the ability to detect multiple objects simultaneously.
Wang et al. [35]	A saliency adjustment mechanism to weight the input image for saliency.	For efficient and accurate detection of ship objects in optical satellite images.
Zhang et al. [36]	Optimizing the network architecture, designing the loss functions, adjusting the prior frames, etc.	Processing large-scale image data for real-time ship inspection.
Zhang et al. [37]	Hybrid attention mechanism of similarity mask.	Efficient and accurate detection of small ship objects in optical remote sensing images.
Zhu et al. [38]	Methods such as introducing attention mechanisms, adapting feature fusion strategies, and optimizing feature propagation paths.	Handling the task of object detection in remote sensing images.
Zhou et al. [39]	A multi-scale feature fusion mechanism is introduced.	Handling vehicle detection tasks and combining multi-scale information to improve detection performance.
Liu et al. [40]	Specific training strategies and optimization methods for aircraft.	Handling the tasks of aircraft inspection and optimizing aircraft-specific problems.
Sharma et al. [41]	Designing a lightweight object detection model.	Handling real-time object detection tasks can be deployed for use on devices with restricted resources or limited computing power.
Wang et al. [42]	Adjusting the network structure, mining, and fusion of ship characteristics.	Handling the task of ship detection in remote sensing images and high detection accuracy can be obtained.

2.1.3. Application of SSD Framework in Remote Sensing Detection

SSD [15] is an efficient single-stage object detection algorithm for object detection tasks in real-time scenarios. It realized object detection at different scales and categories by designing multi-scale feature maps and priori boxes and achieved good performance. In addition to remote sensing images, the SSD algorithm has also been applied to object detection in SAR images [43]. The SSD-based SAR object detection technique improved accuracy and resilience by integrating data augmentation and migration learning, and it also offered fresh suggestions for enhancing the SAR object detection algorithm. In addition, some other methods have achieved specific results in remote sensing image object detection, such as anchor-free methods [44], sparse anchor-guided capsule network method [45], etc. These approaches introduced many technologies, such as adaptive feature selection, channel filter fusion, and attention mechanism [46], to increase the precision and dependability of remote sensing picture object detection. In remote sensing image change detection, Yang et al. [47] proposed an end-to-end deep-learning framework that realized automatic, high-precision, and high-efficiency change detection and had better generalization performance. Additionally, certain enhanced techniques for object detection in remote sensing images were used. For example, the algorithm improved with the attention mechanism and feature fusion achieved excellent performance [48]. When employed to detect small objects in remote sensing photos, the SSD network with dilated convolution and feature fusion [49] produced great results. Meanwhile, Suidong et al. [50] proposed a multi-scale feature fusion and adaptive positive and negative sample filtering mechanism to improve the detection accuracy and performance of small objects in remote sensing images [51]. In conclusion, the SSD algorithm and its upgraded approaches for object detection in remote sensing photos have obtained good performance in many circumstances [52–55], which is of considerable significance and wide application value. However, further research is still needed to solve the specific challenges in remote sensing images, such as the low detection accuracy of small objects, remote objects, and low-contrast objects, in order to improve the performance and robustness of the algorithm.

2.2. Remote Sensing Object Detection Algorithm Based on Candidate Box and Regional Convolutional Neural Network

In the field of object detection, the RCNN series [56–58] algorithms dominated before the advent of YOLOv1. The architecture of the optical remote sensing object recognition algorithm based on the candidate region is shown in Figure 3. The spatially oriented object detection framework that Yu et al. [59] developed had the highest accuracy and resilience in remote sensing images and showed significant application potential in remote sensing image analysis. While this was going on, the upgraded Mask-RCNN model [60] was used to recognize objects in remote sensing images, and the segmentation accuracy and precision were enhanced using new techniques for multi-scale feature fusion, region-generating networks, and segmentation precision optimization. In addition, the aircraft detection method based on multi-scale Faster-RCNN successfully dealt with complex scenes and images of different scales by introducing multi-scale feature pyramids and RoI (region of interest, RoI) complete utilization strategy [61] and achieved excellent detection performance. Sha et al. [62] proposed a multi-scale aircraft object detection method for remote sensing images that improved object detection performance and had significant applications in the fields of aviation monitoring, military investigation, and other areas. In addition, using a multi-scale feature pyramid network and an adaptive label mapping strategy [63], In low-altitude UAV images, Singh et al. [64] proposed an efficient instance segmentation method for railroad sleepers, which was based on the Mask RCNN model and achieved accurate localization and segmentation with high accuracy and efficiency via pixel-level segmentation and optimization strategies. Further, He et al. [65] proposed a seismic wave P-wave arrival pickup method based on the Faster-RCNN model, which improved the picking accuracy by detecting the local window via the object orientation. This way, it can accurately detect the P-wave arrival time and has high computational efficiency. Regarding small object detection in optical remote sensing video, Feng et al. [66] proposed a motion-guided RCNN-based method that achieved high accuracy and robustness by combining motion and appearance features, providing an effective solution for small object detection. In summary, the RCNN series algorithms had a dominant position until the emergence of YOLOv1 and have been widely used in remote sensing image object detection. These methods include candidate region-based algorithms, spatial directional object detection frameworks, and improved Mask-RCNN and Faster-RCNN models. They improve the accuracy, robustness, and segmentation accuracy of object detection by introducing spatial context, direction information, multi-scale feature fusion, and region of interest strategy. However, these methods may face limitations such as high computational complexity and limited detection performance for small objects and complex scenes and need to be further improved and optimized to meet specific application requirements.

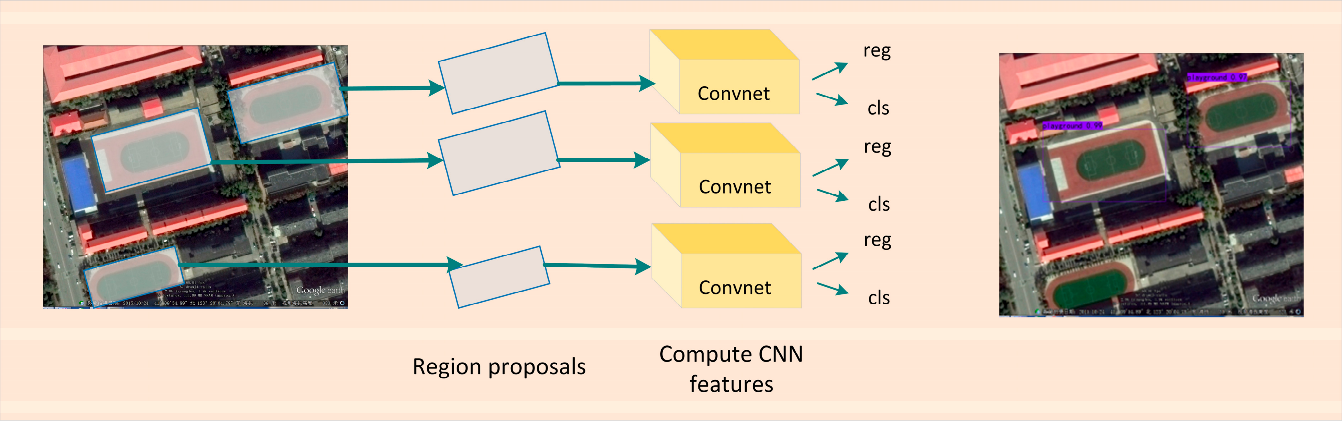

Figure 3. Algorithm flow of optical remote sensing object detection based on candidate region (The architecture diagram includes input and output, region proposal network, compute CNN features, classification, and regression).

2.3. End-to-End Remote Sensing Object Detection Algorithm Based on Transformer Network

The end-to-end remote sensing object detection algorithm based on the Transformer network is an emerging method in the field of object detection and shows potential advantages in remote sensing images. These algorithms extract global context information by introducing a Transformer encoder and using a self-attention mechanism for object detection and object relationship modeling. Some specific methods include a multi-scale visual Transformer based on angle segmentation in Reference [67] and Efficient-Matching-Oriented Object Detection Transformer (EMO2-DETR) in Reference [68]. These methods can handle objects of different scales and directions and achieve high accuracy and robustness in remote sensing images. Compared with the traditional two-stage method, the end-to-end algorithm based on Transformer has the advantages of reducing computational complexity, processing scale changes and complex backgrounds, and providing more global context information, which is conducive to accurate positioning and classification of objects. However, these algorithms still face some challenges. Firstly, transformer-based algorithms usually have high computational complexity and require more computing resources. Secondly, due to the need for a large amount of training data to train the Transformer model, the data demand is enormous, especially in the field of remote sensing images.

Therefore, future research needs to explore further and improve the end-to-end remote sensing object detection algorithm based on Transformer to improve its performance and meet the practical application requirements. This includes research on reducing computational complexity, optimizing the training process, and introducing more domain-specific technologies. In general, the end-to-end remote sensing object detection algorithm based on the Transformer network has broad development prospects and application potential, but it still needs further improvement and research. As shown in Figure 4, the process of optical remote sensing object detection algorithm is based on a Vision Transformer (ViT).

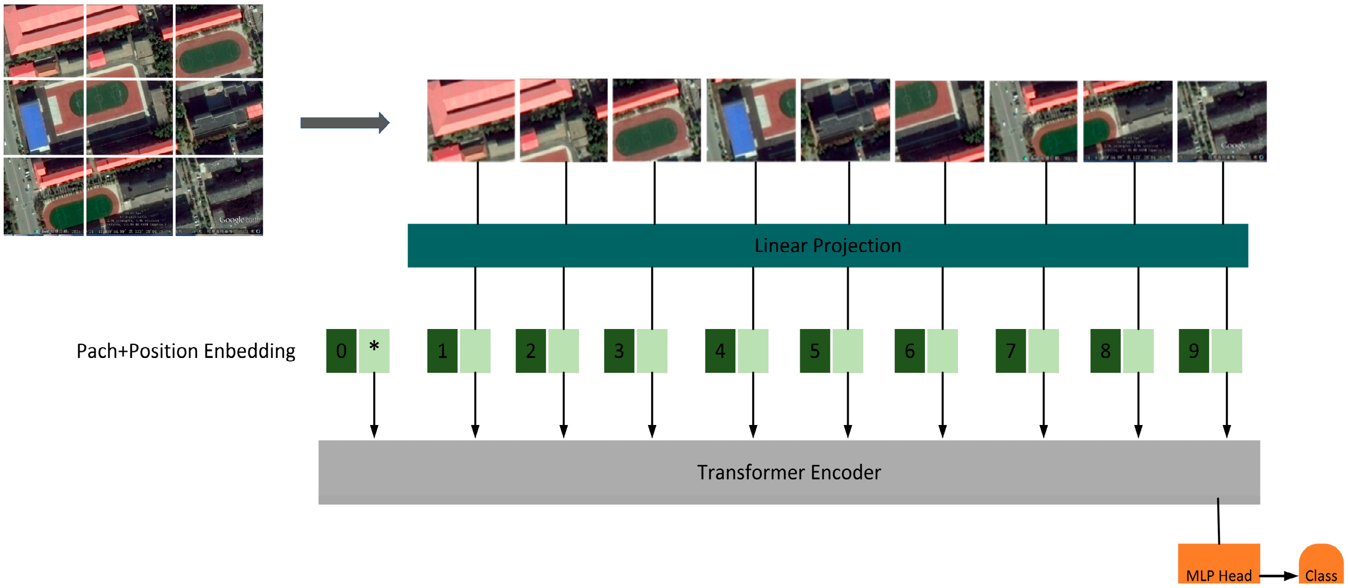

Figure 4. Optical remote sensing object detection algorithm flow based on ViT (the architecture diagram includes input layer, patch embedding, positional embedding, Transformer encoder, global average pooling, and fully connected layer).

2.4. Remote Sensing Object Detection Method for Specific Scenes

2.4.1. Object Detection in Remote Sensing Images Based on Supervision

Deep learning and other technical methods can be used to detect objects in remote sensing photos accurately, and this technique has potential applications in many different fields. For example, Li et al. [69] used deep-learning techniques to detect airports in a real remote sensing environment, extracted features via convolutional neural networks, and used multi-layer perceptrons for classification. With the help of transfer learning and data enhancement technology, satisfactory performance has been achieved. Wu et al. [70] proposed an end-to-end multi-side output fusion deep supervised network for change detection of remote sensing images, which achieved automated and accurate detection with high accuracy and consistency. In addition, there were several studies that used optimized data processing and feature extraction methods [71] to efficiently and accurately extract information from large-scale remote sensing data, which had potential value for remote sensing applications. Meanwhile, some studies based on weakly supervised learning and hierarchical fusion techniques [72] can effectively detect various types of objects and achieve good performance in complex contexts and backgrounds. In addition, the methods based on semi-supervised learning and object-first mixup techniques [73] have also been applied to remote sensing image object detection, which can effectively detect and locate objects with high accuracy and robustness. These studies also explored the application of semi-supervised learning in specific domains, such as power transmission tower object detection and SAR image object detection [74]. These techniques have a wide range of applications and can enhance the effectiveness of object detection by combining both a small amount of labeled data and a large amount of unlabeled data. Among them, the method of introducing semi-supervised migration learning and multi-layer feature fusion mechanisms [75] showed better performance in infrared object detection tasks. In contrast, the SAR object detection network method based on semi-supervised learning [76] improved the detection performance using unlabeled SAR image data to assist training. Finally, the semi-supervised SAR ship object detection method based on a graph attention network had better performance and robustness in ship object detection by constructing a graph structure and applying an attention mechanism [77]. Important application solutions for the fields of marine monitoring and safety prevention were supplied by these studies. The research shows that the method based on deep learning has achieved satisfactory performance in the fields of airport detection, change detection, and information extraction and shows the characteristics of high accuracy, robustness, and automation. However, remote sensing image object detection still faces some challenges, such as high computational complexity, processing requirements for large-scale data, and performance in complex backgrounds. Therefore, future research needs to improve further and optimize the algorithm to improve efficiency, accuracy, and adaptability to meet the needs of practical applications.

2.4.2. Remote Sensing Image Object Detection Method Based on Attention Mechanism

Performance can be increased using the attention mechanism in remote sensing image object detection. It improves the accuracy and robustness of object detection by assigning different attention weights to different regions in remote sensing images so that the network can pay more attention to the object region. Numerous fields, including remote sensing picture processing, geological research, agricultural monitoring, and others can use this technique. Wang et al. [78] proposed an improved model of Double U-Net for remote sensing image change detection. The model adopted a double-head structure and utilized multi-scale feature fusion and separable convolution techniques to efficiently process the images of the current and previous moments to achieve efficient change detection. The experiment proves that the Double U-Net model is superior to the traditional U-Net model with better performance and anti-interference ability, which is expected to be widely promoted in practical applications. A cross-modal attention feature fusion method for object detection in multispectral remote sensing images was proposed by Qingyun et al. [79]. In order to increase the reliability and accuracy of object identification, the technique combined a feature fusion network with a cross-modal attention mechanism. Experiments demonstrated that the method performed well in multispectral remote sensing images and could accurately detect and locate objects, which was of potential application. A quick self-attentive cascade network method for object detection in vast scene remote sensing photos was proposed by Hua et al. [80]. The method utilized the self-attention mechanism to extract features by combining cascade modules and performed object prediction via a classification regression module. Experiments proved that the method had good performance and high detection speed in large scene remote sensing image object detection. A dual-branch network approach for remote sensing image change detection was proposed by Ma et al. [81]. By considering the information and self-attention mechanism of the two moments at the same time, it can accurately detect the changing area and have application value. Zhang et al. [82] proposed a global context detection model using an attention mechanism and multi-scale feature fusion for optical remote sensing image object detection. By concentrating on region formation and feature fusion, our technique enhanced the detection efficiency and accuracy and completed the mission successfully. Nong et al. [83] proposed a method that utilized graph convolutional neural networks and attention mechanisms to achieve spatial relationship detection of remote sensing objects. The method's ability to efficiently gather object connection information and enhance the object detection process's robustness and accuracy was crucial in this field. An attention-guided twin network for change detection in high-resolution remote sensing images was proposed by Yin et al. [84]

The application of attention mechanisms in remote sensing image object detection has shown the potential to improve performance and superior performance in many application fields. In particular, by introducing attention mechanisms, such as cross-modal attention, self-attention, and global context attention, the detection accuracy and robustness of the object can be enhanced. However, the application of attention mechanisms in remote sensing image object detection still faces challenges, such as high computational complexity, processing requirements for large-scale data, and performance optimization in complex scenes. Therefore, future research needs to improve further and optimize the design of attention mechanisms to enhance efficiency, stability, and adaptability to meet practical applications' needs better.

2.4.3. Remote Sensing Image Object Detection Method Based on Multi-Scale Processing

By processing remote sensing images at various scales, the multi-scale remote sensing image object detection approach can increase the detection accuracy and rate. In the specific study, Han et al. [85] proposed a contextual scale-aware detector that enhanced the perception of small and weak objects by introducing context information and scale-aware modules. Dong et al. [86] proposed a multi-scale building detection method based on boundary protection, which effectively solved the problems of missed and false detection in traditional methods. A quick and accurate approach to object detection in remote sensing images was proposed by Zhang et al. [87] and is based on aerial optical sensors. Song et al. [88] proposed an ERMF (edge refinement multi-feature, ERMF) method to improve the accuracy of dual-temporal remote sensing image change detection by fusing multiple features and adopting a refined segmentation strategy. Gao et al. [89] proposed a global-to-local (GL) network, which achieved accurate and robust remote sensing object detection via global and local feature fusion. Chen et al. [90] proposed the Info-FPN network improved the remote sensing image object detection algorithm via an information transfer mechanism and an adaptive weighted loss function. Su et al. [91] proposed the multi-scale context-aware RCNN method, which showed high accuracy and stability in detecting small sample objects in remote sensing images. Dong et al. [92] proposed the multi-scale deformable attention and multilevel feature aggregation method, which exhibited high accuracy and robustness in remote sensing object detection. Zhang et al. [93] proposed the multi-scale structural conditional feature transformation network showed high accuracy and robustness in object detection of remote sensing images. Dong et al. [94] proposed the convolutional neural network method based on appropriate object scale features to solve the detection problem caused by object scale differences in high-resolution remote sensing images using multi-scale features and adaptive pooling operations. Meng et al. [95] proposed a multi-scale convolutional neural network remote sensing object detection method that demonstrated higher accuracy and robustness on real datasets. Yao et al. [96] proposed the remote sensing multi-scale object detection method that utilized multivariate feature extraction and characterization optimization to achieve efficient and accurate detection of different scale objects in remote sensing images. Zhang et al. [97] proposed an all-around accurate detection algorithm for dense small objects and achieved efficient and accurate detection by introducing RPN and multi-scale feature fusion. Zhang et al. [98] proposed an image segmentation method that utilized an adaptive pyramid network and a parallel spatial channel to enhance performance. Zhou et al. [99] proposed the APS-Net (adaptive point set network, APS-Net), which achieved an accurate detection of complex scenes and small objects via adaptive mechanisms and multi-scale feature fusion. As shown in Table 3, the remote sensing image object detection method based on multi-scale processing has many applications in improving detection accuracy and detection rate. These methods effectively solve the problems of object size change and background complexity in remote sensing images by introducing multi-scale information, context awareness, feature fusion, and other technologies, and achieve good performance. However, these methods may face the challenges of high computational complexity and extensive memory usage when dealing with large-scale data and may have poor detection results for objects with large-scale changes. Therefore, in practical applications, it is necessary to optimize the algorithm's efficiency and robustness to meet the needs of large-scale and multi-scale remote sensing image object detection.

Table 3. Comparison of object detection algorithms for remote sensing images based on multi-scale.

Literature	Paper Highlights	Applicability
Han et al. [85]	Context scale-aware detector, providing a new benchmark data set.	Remote sensing small weak object detection task in UAV images.
Dong et al. [86]	Accurate maintenance of building boundaries is emphasized, and the accuracy and robustness of detection are improved using a multi-scale strategy.	Building detection tasks in earthquake disasters and other remote sensing images.
Zhang et al. [87]	An efficient object detection method based on multi-scale aerial optical sensor.	Remote sensing image, aerial photography, UAV image analysis, and other fields.
Song et al. [88]	Innovative design with edge refinement and multi-feature fusion.	Two-phase remote sensing image change detection, surface environmental change monitoring, and other tasks.
Gao et al. [89]	Innovative design with scale perception and global-to-local strategy.	Remote sensing object detection and multi-scale object detection tasks.
Chen et al. [90]	An innovative design with an information feature pyramid and feature selection based on information gain.	
Su et al. [91]	The small sample object detection of multi-scale context-aware.	Few-shot object detection task in remote sensing images
Dong et al. [92]	Multi-scale deformable attention mechanism and multi-level feature aggregation method are used to improve the accuracy and robustness of object detection.	Remote sensing image object detection tasks.
Zhang et al. [93]	The multi-scale structural condition feature transformation and attention module are introduced.	
Dong et al. [94]	The method of applying object scale feature extraction and structural optimization.	High-resolution remote sensing image object detection task.
Meng et al. [95]	The method of multivariate feature extraction and characterization optimization.	Remote sensing multi-scale object detection tasks.
Yao [96]	The method of multi-scale fusion feature and convolutional neural network	Remote sensing imagery aircraft object detection Mission.
Zhang [97]	Using the specific object detection network structure and algorithm optimization technology to achieve accurate detection of dense small objects.	Detection task of dense small objects in remote sensing images.
Zhang [98]	An adaptive point set network and a point set modeling and matching method are introduced.	Optical remote sensing image target detection task.
Zhou [99]	Autonomous structure pyramid network and parallel space-channel attention mechanism.	Change detection task of high-resolution remote sensing images.

2.4.4. Based on Deep Learning and Traditional Manual Feature Extraction Methods

Deep-learning technology has made significant progress in the field of computer vision and pattern recognition. However, traditional manual feature extraction methods are still significant in some fields. A deep-learning network structure was proposed by Wang et al. [100]. It is based on two-dimensional discrete wavelet transform and adaptive feature weighted fusion. Xu et al. [101] proposed a new neural network structure that realized pixel-by-pixel classification of high-resolution remote sensing images by introducing control gates and feedback attention mechanisms. Liu et al. [102] proposed a new network architecture called contourlet CNN (C-CNN). It combines traditional spectral analysis with convolutional neural networks to make feature representation for texture classification tasks sparse and effective. Hu et al. [103] proposed an object detection method based on a small number of samples called Gabor-CNN. By constructing a feature extraction convolution kernel library, improving the region proposal process, and using a deep utilization feature pyramid network, the accuracy and recall rate on a small sample data set are better than the existing comparison models, which have a strong application prospect. Chen et al. [104] proposed a hyperspectral image classification method that combines the Gabor filter and convolution filter. This would help alleviate the overfitting problem of deep CNN when there are not enough training samples. Zheng et al. [105] proposed a new frequency domain direction learning module that encodes the direction information of directional object detection via a frequency domain feature extraction network and a direction-enhanced self-attention layer. El-Khamy et al. [106] proposed a new convolutional neural network model using a discrete wavelet transform pool (DWTPL) to extract spectral information. They obtained a better classification accuracy and f1 score than other models on the AID dataset. This method can effectively process high-resolution images and improve scene classification performance. EL GAYAR et al. [107] proposed a new image decoding model, combining a wavelet-driven convolutional neural network with a two-stage discrete wavelet transform to extract salient features in images. Using the deep visual prediction model and long-term and short-term memory as the decoder, the model can automatically generate semantic titles according to the image's semantic context information and spatial features. He et al. [108] proposed a method combining the adequate channel attention (ECA) module with the ResNet model to improve the accuracy of remote sensing image classification. The ECA module shows high accuracy on the AID dataset by avoiding dimensionality reduction and using a deep residual structure. Tsourounis et al. [109] proposed a new SIFT-CNN combination method, which implicitly obtains local rotation invariance by converting SIFT descriptors into multi-channel images and training CNN to use these images as input. Li et al. [110] proposed a color texture convolutional neural network, which combines the characteristics of traditional CNN and Gabor CNN. The image classification task is completed by inputting the original image and the Gabor processed image into two stream networks and combining the output at the end.

In summary, based on some research results of deep learning and traditional methods, these methods improved the accuracy and efficiency of remote sensing image processing. However, these methods may require a lot of data support and complex computing resources, and there are still limitations in specific scenarios, which need further practical application and improvement.

2.4.5. Fast Image Processing Method Based on VHR

With the wide application of VHR images, it is necessary to develop fast processing methods to improve processing efficiency. Huo et al. [111] proposed a new high-resolution image change detection method that uses fast object-level feature extraction and progressive change feature classification. By improving the distinguishability between changed classes and unchanged classes, dynamically adjusting training samples, and gradually adjusting the separation hyperplane, this method achieves higher classification accuracy and automation, and experiments prove its effectiveness. Mboga et al. [112] developed a land cover classification method based on a fully convolutional network, using aerial RGB images for end-to-end training and introducing skip connections to restore high spatial details. This method has a lower computational cost and performs better in edge depiction. Zhang et al. [113] proposed a coarse-to-fine large-size ultra-high resolution image registration method, which accelerates the acquisition of control points and image correction by calculating a unified device architecture. Saha et al. [114] proposed an unsupervised context-sensitive framework, called depth change vector analysis, for change detection of multi-temporal high spatial resolution images. This method uses convolutional neural network features to model the spatial relationship between pixels. It uses a hierarchical automatic feature selection strategy to select features emphasizing the change information of high and low prior probabilities. The depth change vector is obtained by comparing the depth feature supervector, and then the change pixels are identified. The experimental results show that the proposed method achieves effective change detection results on different data sets.

In summary, some fast for high-resolution remote sensing image processing methods are introduced. These methods have achieved specific results in improving processing efficiency and accuracy. However, in practical applications, it is still necessary to pay attention to the robustness and generalization ability of these methods in different scenarios, and further verify their applicability in large-scale data and complex environments.

3. Performance Evaluation and Comparison of Optical Remote Sensing Image Object Detection

3.1. Optical Remote Sensing Image Data Sets

Optical remote sensing image data sets play a vital role in remote sensing object detection tasks. These datasets provide valuable standard remote sensing data for model training and an objective and unified benchmark for comparing different networks and algorithms. In recent years, with the development of satellite remote sensing technology, some high-quality optical remote sensing image object detection data sets have become increasingly popular. In this paper, 15 representative data sets are selected for introduction. Their sample statistical information is shown in Table 4, including publisher, and content description, number of object categories, and number of images contained in the data set. These open optical remote sensing image data sets promote the rapid development of remote sensing object detection technology based on deep learning.

Table 4. Overview of commonly used optical remote sensing image object detection datasets [115].

Data Set	Publisher and Content Description	Number of Object Categories	Number of Images
TAS [116]	Vehicle objecting dataset published by Stanford University.	1	30
OIRDS [117]	Vehicle objecting datasets published by Raytheon Corporation.	5	900
SZTAKI [118]	Rotating building object dataset published by Mta Sztaki.	1	9
UCAS-AOD [119]	Vehicle and aircraft object datasets published by CAS, and background negative samples.	2	976
NWPU VHR-10 [120]	The data set of aircraft, ships, oil tanks, baseball courts, tennis courts, basketball courts, and other objects released by Northwestern Polytechnical University.	10	1510
VEDAI [121]	Vehicle objecting dataset published by Caen University.	9	1210
HRSC2016 [122]	Ship objecting dataset released by Northwestern Polytechnical University.	1	1061
DLR3k [123]	Vehicle object dataset published by German Aerospace Center.	7	20
RSOD [124]	Aircraft, oil tank, stadium, and overpass object datasets released by Wuhan University.	4	976
TGRS-HRRSD [125]	Object datasets for ships, bridges, athletic fields, oil tanks, basketball courts, tennis courts, and other object data sets released by the Chinese Academy of Sciences.	13	21,761

Table 4. *Cont.*

Data Set	Publisher and Content Description	Number of Object Categories	Number of Images
LEVIR [126]	The object data set of aircraft, ships, and oil tanks released by Beijing University of Aeronautics and Astronautics.	3	22,000
ITCVD [127]	Vehicle objecting dataset published by Twente University.	1	135
DIOR [128]	Aircraft, airports, basketball courts, bridges, chimneys, dams, and other object data sets published by Northwestern Polytechnical University.	20	23,463
DOTA [129]	object data sets of ships, swimming pools, track and field fields, ports, helicopters, football fields, and other object data sets released by Wuhan University.	16	2806
FAIR1M [130]	The data set of 5 large categories and 37 fine-grained categories such as aircraft, ships, vehicles, stadiums, and roads published by the Chinese Academy of Sciences is the world's largest fine-grained object detection and recognition data set for optical remote sensing images.	37	15,000

3.2. Algorithm Performance Evaluation and Comparison

Currently, the commonly used performance indicators for evaluating optical remote sensing image object detection algorithms are precision, recall, mean average precision (mAP), and frame per second (FPS). The accuracy reflects the proportion of actual positive samples in the test results. The recall rate reflects the proportion of positive samples that are correctly detected in all positive samples to be noticed, and there is a trade-off combination between accuracy and recall rate. The precision–recall curve (PR) can be obtained by plotting the precision as the ordinate and the recall as the abscissa. The area under the curve represents the AP of a specific category of objects. The AP mean of multiple categories is the average precision mean mAP, which represents the algorithm's overall performance on the data set. Table 5 displays a performance comparison of popular optical remote sensing picture object detection techniques. The following conclusions may be drawn from the previous analysis of the critical properties of various algorithms and the performance comparison of typical algorithms on the same data set in Table 5.

Table 5. Performance comparison of typical optical remote sensing image object detection algorithms.

Data Set	Algorithm	Backbone Network	Literature	Release Time	mAP/%
RSOD [124]	FPN-YOLO	DarkNet53	Sun et al. [28]	2021	87.40
	DAM-YOLOX	CSPDarkNet	Wei et al. [29]	2023	93.90
	YOLOv5-DNA		Xin et al. [34]	2022	77.51
	DF-SSD	ResNet-50	Qu et al. [45]	2020	51.78
	SSOD-RS		Zhang et al. [73]	2021	90.70
	RCNN-FCD		Su et al. [91]	2022	96.60
	MLFAM		Dong et al. [92]	2022	92.50
	I-SSD	VGG-16	Liu et al. [47]	2022	80.53
	AFF-SSD		Yin et al. [55]	2022	75.19

Table 5. *Cont.*

Data Set	Algorithm	Backbone Network	Literature	Release Time	mAP/%
DOTA [129]	GSC-YOLO	CSPDarkNet	Ma et al. [30]	2022	93.44
	YOLOv4-CD		Zhu et al. [38]	2023	90.88
	SAHR-CapsNet		Yu et al. [50]	2021	93.04
	AF-SSD	ResNet-50	Lu et al. [44]	2021	52.60
	FFC-SSD		Xue et al. [52]	2022	74.90
	SOSA-FCN		Hua et al. [80]	2020	95.25
	ATMTransformer	DETR	Zhang et al. [67]	2022	77.30
	EMO2-DETR		Hu et al. [68]	2023	70.91
	Info-FPN	FPN	Chen et al. [90]	2023	75.84
FAIR1M [130]	YOLM	CSPDarkNet	Liu et al. [40]	2022	88.70
NWPU VHR-10 [120]	AF-SSD	ResNet-50	Lu et al. [44]	2021	69.80
	DF-SSD		Qu et al. [45]	2020	65.35
	FESSD		Shi et al. [54]	2020	79.36
	MSCNN		Yao et al. [96]	2019	96.00
	CenterNet	DLA-34	Liu et al. [49]	2020	95.70
	GCDN	ResNet-18	Zhang et al. [82]	2020	97.60
DIOR [128]	RSADet	DLA-34	Yu et al. [59]	2021	72.20
	CenterNet		Yu et al. [59]	2021	69.40
DIOR [128]	MLFAM	ResNet-50	Dong et al. [92]	2022	73.90
	MFC	MFE	Meng et al. [95]	2023	70.90
VEDAI [121]	YOLOFusion	CSPDarkNet	Qingyun [79]	2022	78.60
TGRS-HRRSD [125]	SOSA-FCN	ResNet-50	Hua et al. [80]	2020	97.25
	MSFT	SCFT	Zhang et al. [93]	2021	86.33
	MFC	MFE	Meng et al. [95]	2023	90.20

4. Challenge and Improvement Direction

For remote sensing object detection tasks, selecting data sets is crucial. Large-scale, diversified, and high-resolution data sets are needed, and the balance of different categories of samples is maintained. In terms of annotation, the bounding box is usually used to mark the position and size of the object, and the partial annotation or splitting of overlapping objects can be considered. For multi-class object detection, it can be transformed into a multi-label classification problem, which combines prior knowledge and deep-learning models to improve accuracy. For minor sample problems, data augmentation techniques and transfer learning methods can be used. In order to improve robustness, data enhancement, abnormal sample detection and repair, multi-scale information fusion, and other measures can be taken. To improve the interpretability of the model, many methods can be employed, including visual attention, model anatomy, introduction of rules, and interpretable network architecture.

When employing deep-learning techniques for remote sensing object detection, it is imperative to consider various challenges, such as computational cost, availability of datasets, and real-time processing capabilities. Many techniques can be employed to reduce the computational cost, including adopting lightweight models, utilizing hardware accelerators, implementing distributed computing, incorporating hardware acceleration libraries, and applying optimization algorithms. When the dataset is limited and expensive, it is possible to find a suitable dataset to cooperate or consider using synthetic data for training and reduce the amount of calculation via data preprocessing and enhancement. In terms of real-time processing capabilities, selecting real-time object detection algorithms and using parallel computing based on GPU clusters can improve processing speed.

5. Conclusions and Prospect

This paper summarizes the research progress of object detection in optical remote sensing images based on the YOLO series, SSD series, candidate region series, and Transformer algorithm. Among them, the YOLO series algorithm has fast detection speed and high accuracy, which is suitable for real-time scenes, but has limitations in small object detection; SSD series algorithms use multi-scale feature maps for object detection, which can better solve the problem of small object detection, but the detection accuracy is relatively low. The candidate region series algorithm performs object detection in two steps: candidate region extraction and classification, which has high accuracy but high computational complexity. The Transformer algorithm has made a significant breakthrough in the field of natural language processing. In recent years, it has also been applied to object detection tasks with good detection performance and interpretability.

Future research needs to focus on the following directions: small object detection, multi-scale information fusion, diversity and scale of data sets, and real-time and efficiency of algorithms. These directions will promote the development and application of optical remote sensing image object detection. In addition, there are some potential areas of concern:

Applying attention to new remote sensing tasks: In addition to object detection, the attention mechanism can be applied to other remote sensing tasks, such as instance segmentation, scene classification, etc. Researchers can explore how to use attention mechanisms to improve the performance and efficiency of these tasks.

Developing interpretable attention mechanisms: In order to improve the model interpretability, researchers can explore the development of interpretable attention mechanisms to explain model decisions. This can enhance the user’s trust in model prediction and help explain the model’s attention to different objects in remote sensing images.

Attention for multi-modal and multi-source remote sensing data fusion: With the progress of remote sensing technology, multi-modal and multi-source remote sensing data have become more abundant. Researchers can study how to effectively integrate information from multi-modal data (such as optics, radar, hyperspectral, etc.) and multi-source data (such as satellites, aircraft, ground sensors, etc.) to improve the performance and robustness of object detection.

Attention mechanisms for few-shot and semi-supervised learning with limited labeled data: In optical remote sensing image object detection, obtaining a large amount of labeled data may be difficult and expensive. Therefore, researchers can explore the semi-supervised learning method with few samples and combine the attention mechanism to make full use of limited labeled data to improve the model’s performance.

Lightweight attention modules to reduce computational complexity: In order to achieve real-time and efficiency in a resource-constrained environment, researchers can develop lightweight attention modules to reduce computational complexity. This will make it possible to perform optical remote sensing image object detection in an environment with limited embedded devices or computing resources.

Self-supervised pre-training with attention for improved generalization: Self-supervised learning has shown potential in optical remote sensing image object detection. Further research can explore how to combine the attention mechanism for self-supervised pre-training to improve the model’s generalization performance in different datasets and scenarios.

Extending to video analysis and change detection over time: In addition to static images, researchers can extend the attention mechanism to the field of remote sensing video analysis, such as object tracking, behavior recognition, and time-varying change detection. This will help to better understand the dynamic scene in remote sensing images and provide more comprehensive information.

Author Contributions: X.B. has made significant contributions to the techniques or methods used and research concepts of the articles. K.W. has made significant contributions to the critical revision of the articles. C.B. has made significant contributions to the data collection and design of the articles. The final draft was read and approved by all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation Key Project of Gansu Province (Grant No. 23JRRA860), the Natural Science Foundation of Gansu Province (Grant No. 23JRRA913) and the key talent project of Gansu Province. Key R & D and achievement transformation project of Inner Mongolia Autonomous Region (2023YFSH0043,2023YFDZ0043).

Data Availability Statement: Not Applicable.

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

References

1. Cheng, G.; Han, J.W. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]
2. Chen, Q.; Tang, S.; Yang, Q.; Fu, S. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In Proceedings of the 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), Dallas, TX, USA, 7–9 July 2019; IEEE: Piscataway Township, NJ, USA, 2019; pp. 514–524.
3. Najibi, M.; Samangouei, P.; Chellappa, R.; Davis, L.S. Ssh: Single stage headless face detector. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4875–4884.
4. Zhang, L.; Lin, L.; Liang, X.; He, K. Is faster RCNN doing well for pedestrian detection? In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 443–457.
5. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
6. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef] [PubMed]
7. Vedantam, R.; Lawrence Zitnick, C.; Parikh, D. Cider: Consensus-based image description evaluation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4566–4575.
8. Reddy, K.R.; Priya, K.H.; Neelima, N. Object Detection and Tracking—A Survey. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; IEEE: Piscataway Township, NJ, USA, 2015; pp. 418–421.
9. Kuehne, H.; Jhuang, H.; Garrote, E.; Poggio, T.; Serre, T. HMDB: A large video database for human motion recognition. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway Township, NJ, USA, 2011; pp. 2556–2563.
10. Lowe, G.D. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
11. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; IEEE Computer Society: Washington, DC, USA, 2005; pp. 886–893.
12. Felzenszwalb, P.; McAllester, D.; Ramanan, D. A discriminatively trained, multiscale, deformable part model. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; IEEE: Piscataway Township, NJ, USA, 2008; pp. 1–8.
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster RCNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*.
14. 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.
15. 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 Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
16. Redomn, J.; Farhadi, A. YOLO9000: Better, faster, stron-ger. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Honolulu, HW, USA, 21–26 July 2017; 1EEE: Piscataway Township, NJ, USA, 2017.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
19. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
20. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.

21. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
22. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
23. Li, Y.; Fan, Q.; Huang, H.; Han, Z.; Gu, Q. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. *Drones* **2023**, *7*, 304. [CrossRef]
24. Tang, X.; Zhang, X.; Shi, J.; Wei, S. A moving object detection method based on YOLO for dual-beam SAR. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 5315–5318.
25. Jindal, M.; Raj, N.; Saranya, P.; Sundarabalan, V. Aircraft Detection from Remote Sensing Images using YOLOV5 Architecture. In Proceedings of the 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 21–22 April 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 332–336.
26. Shi, Y. An Underwater Target Wake Detection in Multi-Source Images Based on Improved YOLOv5. *IEEE Access* **2023**, *11*, 31990–31996. [CrossRef]
27. Ding, W.; Zhang, L. Building detection in remote sensing image based on improved YOLOv5. In Proceedings of the 2021 17th International Conference on Computational Intelligence and Security (CIS), Chengdu, China, 19–22 November 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 133–136.
28. Sun, Y.; Liu, W.; Hou, X.; Bi, F. FRN-YOLO: A Feature Re-fusion Network for Remote Sensing object detection. In Proceedings of the 2021 2nd International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China, 12–14 November 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 372–375.
29. Wei, J.; Liu, Y.; Li, L.; Xie, W.; Zhao, S.; Zhao, Z. Improved YOLO X with Bilateral Attention for Small Object Detection. In Proceedings of the 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC), Zakopane, Poland, 16–17 June 2023; IEEE: Piscataway Township, NJ, USA, 2023; pp. 1–6.
30. Ma, L.; He, T.; Sun, Y.; Hu, B.B. Lightweight YOLOv4 Algorithm for Remote Sensing Image Detection. In Proceedings of the 2022 14th International Conference on Signal Processing Systems (ICSPS), Jiangsu, China, 18–20 November 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 793–797.
31. Li, X.; Cai, K. Method research on ship detection in remote sensing image based on Yolo algorithm. In Proceedings of the 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Xi'an, China, 14–16 August 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 104–108.
32. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6083–6101. [CrossRef]
33. Yang, Y.; Liao, Y.; Cheng, L.; Zhang, K.; Wang, H.; Chen, S. Remote sensing image aircraft object detection based on giou-yolo v3. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 474–478.
34. Xin, L.; Xie, X.; Lv, J. Research on Remote Sensing Image object detection Algorithm Based on YOLOv5. In Proceedings of the 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 16–18 December 2022; IEEE: Piscataway Township, NJ, USA, 2022; Volume 5, pp. 1497–1501.
35. Wang, S.; Sun, H.; Zhu, Y.; Li, M.; Xu, Q. SA-YOLO: The Saliency Adjusted Deep Network for Optical Satellite Image Ship Detection. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 2131–2134.
36. Zhang, X.; Zhang, Z. Ship detection based on improved YOLO algorithm. In Proceedings of the 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 5–8 January 2023; IEEE: Piscataway Township, NJ, USA, 2023; pp. 98–101.
37. Zhang, X.; Yuan, S.; Luan, F.; Lv, J.; Liu, G. Similarity Mask Mixed Attention for YOLOv5 Small Ship Detection of Optical Remote Sensing Images. In Proceedings of the 2022 WRC Symposium on Advanced Robotics and Automation (WRC SARA), Beijing, China, 20 August 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 263–268.
38. Zhu, F.; Wang, Y.; Cui, J.; Liu, G.; Li, H. object detection for remote sensing based on the enhanced YOLOv4 with improved BiFPN. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 351–360.
39. Zhou, L.; Liu, J.; Chen, L. Vehicle detection based on remote sensing image of Yolov3. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; IEEE: Piscataway Township, NJ, USA, 2020; Volume 1, pp. 468–472.
40. Liu, W.; Tian, J.; Tian, T. YOLM: A Remote Sensing Aircraft Detection Model. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 1708–1711.
41. Sharma, M.; Markopoulos, P.P.; Saber, E. YOLOrs-lite: A lightweight cnn for real-time object detection in remote-sensing. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 2604–2607.

42. Wang, Y.K.; Jiang, H.X.; Lin, K.Y. Remote sensing image ship detection based on modified YOLO algorithm. *J. Beijing Univ. Aeronaut. Astronaut.* **2020**, *46*, 1184–1191. [CrossRef]
43. Wang, Z.; Du, L.; Mao, J.; Liu, B.; Yang, D. SAR object detection based on SSD with data augmentation and transfer learning. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 150–154. [CrossRef]
44. Liu, Y.; Yang, J.; Cui, W. Simple, Fast, Accurate Object Detection based on Anchor-Free Method for High Resolution Remote Sensing Images. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 2443–2446.
45. Yu, Y.; Wang, J.; Qiang, H.; Jiang, M.; Tang, E.; Yu, C.; Zhang, Y.; Li, J. Sparse anchoring guided high-resolution capsule network for geospatial object detection from remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102548. [CrossRef]
46. Wang, Z.H.; Yang, C. Improved SSD model in extraction application of expressway toll station locations from GaoFen 2 remote sensing image. *J. Traffic Transp. Eng.* **2021**, *21*, 278–286. [CrossRef]
47. Yang, Y.; Gu, H.; Han, Y.; Li, H. An end-to-end deep learning change detection framework for remote sensing images. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 652–655.
48. Lu, X.; Ji, J.; Xing, Z.; Miao, Q. Attention and feature fusion SSD for remote sensing object detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–9. [CrossRef]
49. Qu, J.; Su, C.; Zhang, Z.; Razi, A. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images. *IEEE Access* **2020**, *8*, 82832–82843. [CrossRef]
50. Suidong, L.; Lei, Z.H.U.; Wenwu, W. Improving SSD for detecting small target in Remote Sensing Image. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 567–571.
51. Liu, S.; Shi, H.; Guo, Z. Remote sensing image object detection based on improved SSD. In Proceedings of the 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), Changchun, China, 20–22 May 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 421–424.
52. Han, J.; Zhang, D.; Cheng, G.; Guo, L.; Ren, J. Object Detection in Optical Remote Sensing Images Basedon FFC-SSD Model. *Acta Opt. Sin.* **2022**, *42*, 138–148.
53. Wang, H.T. Research on Target Detection Algorithm of Optical Aircraft Remote Sensing Image Based on Improved SSD. Master's Thesis, Ningxia University, Yinchuan, China, 2022. [CrossRef]
54. Shi, W.X.; Tan, D.L.; Bao, S.L. Feature Enhancement SSD Algorithm and Its Application in Remote Sensing Images Target Detection. *Acta Photonica Sin.* **2020**, *49*, 154–163.
55. Yin, F.L.; Wang, T.Y. Target detection of remote sensing image based on attention feature fusion SSD algorithm. *Netw. Secur. Data Gov.* **2022**, *41*, 67–73. [CrossRef]
56. 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, 23–28 June 2014; pp. 580–587.
57. Girshick, R. Fast RCNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
58. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask RCNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
59. Yu, D.; Ji, S. A new spatial-oriented object detection framework for remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
60. Huiming, Y.; Fuxin, X. A remote sensing image target recognition method based on improved Mask-RCNN model. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 March 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 436–439.
61. Miao, W.X.; Luo, Z. Aircraft detection based on multiple scale faster-RCNN. In Proceedings of the 2018 International Conference on Virtual Reality and Visualization (ICVRV), Qingdao, China, 22–24 October 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 90–93.
62. Sha, M.M.; Li, Y.; Li, A. Multiscale aircraft detection in optical remote sensing imagery based on advanced Faster R-CNN. *Natl. Remote Sens. Bull.* **2022**, *26*, 1624–1635. [CrossRef]
63. Wang, Y.; Bashir, S.M.A.; Khan, M.; Ullah, Q.; Wang, R.; Song, Y.; Guo, Z.; Niu, Y. Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Syst. Appl.* **2022**, *197*, 116793. [CrossRef]
64. Singh, A.K.; Dwivedi, A.K.; Sumanth, M.; Singh, D. An efficient approach for instance segmentation of railway track sleepers in low altitude UAV images using mask RCNN. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 4895–4898.
65. He, Z.; Peng, P.; Wang, L.; Jiang, Y. Enhancing seismic p-wave arrival picking by target-oriented detection of the local windows using faster-rcnn. *IEEE Access* **2020**, *8*, 141733–141747. [CrossRef]

66. Feng, J.; Liang, Y.; Ye, Z.; Wu, X.; Zeng, D.; Zhang, X.; Tang, X. Small object detection in optical remote sensing video with motion guided RCNN. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway Township, NJ, USA, 2020; pp. 272–275.
67. Zhang, C.; Liu, T.; Lam, K.M. Angle Tokenization Guided Multi-Scale Vision Transformer for Oriented Object Detection in Remote Sensing Imagery. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 3063–3066.
68. Hu, Z.; Gao, K.; Zhang, X.; Wang, J.; Wang, H.; Yang, Z.; Li, C.; Li, W. EMO2-DETR: Efficient-Matching Oriented Object Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5616814. [CrossRef]
69. Li, N.; Cheng, L.; Ji, C.; Chen, H.; Geng, W.; Yang, W. Airport detection in remote sensing real-open world using deep learning. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106083. [CrossRef]
70. Wu, X.; Yang, L.; Ma, Y.; Wu, C.; Guo, C.; Yan, H.; Qiao, Z.; Yao, S.; Fan, Y. An end-to-end multiple side-outputs fusion deep supervision network based remote sensing image change detection algorithm. *Signal Process.* **2023**, *213*, 109203. [CrossRef]
71. Li, Y.; Li, X.; Zhang, Y.; Peng, D.; Bruzzone, L. Cost-efficient information extraction from massive remote sensing data: When weakly supervised deep learning meets remote sensing big data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103345. [CrossRef]
72. Wu, Z.Z.; Xu, J.; Wang, Y.; Sun, F.; Tan, M.; Weise, T. Hierarchical fusion and divergent activation based weakly supervised learning for object detection from remote sensing images. *Inf. Fusion* **2022**, *80*, 23–43. [CrossRef]
73. Zhang, Z.; Feng, Z.; Yang, S. Semi-supervised object detection framework with object first mixup for remote sensing images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 2596–2599.
74. Zha, W.; Hu, L.; Duan, C.; Li, Y. Semi-supervised learning-based satellite remote sensing object detection method for power transmission towers. *Energy Rep.* **2023**, *9*, 15–27. [CrossRef]
75. Li, W.P.; Yang, X.G.; Li, C.X.; Lu, R.; Huang, P. An improved semi-supervised transfer learning method for infrared object detection neural network. *Infrared Laser Eng.* **2021**, *50*, 243–250.
76. Du, L.; Wei, D.; Li, L.; Guo, Y. SAR Target Detection Network via Semi-supervised Learning. *J. Electron. Inf. Technol.* **2020**, *42*, 154–163.
77. Lv, J.D.; Wang, T.; Tang, X.B. Semi-supervised SAR Ship Target Detection with Graph Attention Network. *J. Electron. Inf. Technol.* **2023**, *45*, 1541–1549.
78. Wang, X.; Yan, X.; Tan, K.; Pan, C.; Ding, J.; Liu, Z.; Dong, X. Double U-Net (W-Net): A change detection network with two heads for remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103456. [CrossRef]
79. Qingyun, F.; Zhaokui, W. Cross-modality attentive feature fusion for object detection in multispectral remote sensing imagery. *Pattern Recognit.* **2022**, *130*, 108786. [CrossRef]
80. Hua, X.; Wang, X.; Rui, T.; Zhang, H.; Wang, D. A fast self-attention cascaded network for object detection in large scene remote sensing images. *Appl. Soft Comput.* **2020**, *94*, 106495. [CrossRef]
81. Ma, C.; Weng, L.; Xia, M.; Lin, H.; Qian, M.; Zhang, Y. Dual-branch network for change detection of remote sensing image. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106324. [CrossRef]
82. Zhang, R.T.; Jiang, X.J.; An, J.S.; Cui, T.S. Design of global-contextual detection model for optical remote sensing targets. *Chin. Opt.* **2020**, *13*, 1302–1313.
83. Nong, Y.J.; Wang, J.J. Spatial Relation ship Detection Method of Remote Sensing Objects. *Acta Opt. Sin.* **2021**, *41*, 212–217.
84. Yin, H.; Weng, L.; Li, Y.; Xia, M.; Hu, K.; Lin, H.; Qian, M. Attention-guided siamese networks for change detection in high resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103206. [CrossRef]
85. Han, W.; Li, J.; Wang, S.; Wang, Y.; Yan, J.; Fan, R.; Zhang, X.; Wang, L. A context-scale-aware detector and a new benchmark for remote sensing small weak object detection in unmanned aerial vehicle images. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102966. [CrossRef]
86. Dong, Z.; Zhang, M.; Li, L.; Liu, Q.; Wen, Q.; Wang, W.; Luo, W.; Wu, Z.; Tang, T.; Ji, W. A multiscale building detection method based on boundary preservation for remote sensing images: Taking the Yangbi M6. 4 earthquake as an example. *Nat. Hazards Res.* **2022**, *2*, 121–131. [CrossRef]
87. Zhang, Q.; Tang, J.; Zheng, H.; Lin, C. Efficient object detection method based on aerial optical sensors for remote sensing. *Displays* **2022**, *75*, 102328. [CrossRef]
88. Song, Z.; Li, X.; Zhu, R.; Wang, Z.; Yang, Y.; Zhang, X. ERMF: Edge refinement multi-feature for change detection in bitemporal remote sensing images. *Signal Process. Image Commun.* **2023**, *2023*, 116964. [CrossRef]
89. Gao, T.; Niu, Q.; Zhang, J.; Chen, T.; Mei, S.; Jubair, A. Global to Local: A Scale-Aware Network for Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5615614. [CrossRef]
90. Chen, S.; Zhao, J.; Zhou, Y.; Wang, H.; Yao, R.; Zhang, L.; Xue, Y. Info-FPN: An Informative Feature Pyramid Network for object detection in remote sensing images. *Expert Syst. Appl.* **2023**, *214*, 119132. [CrossRef]
91. Su, H.; You, Y.; Meng, G. Multi-Scale Context-Aware RCNN for Few-Shot Object Detection in Remote Sensing Images. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 1908–1911.
92. Dong, X.; Qin, Y.; Fu, R.; Gao, Y.; Liu, S.; Ye, Y.; Li, B. Multiscale deformable attention and multilevel features aggregation for remote sensing object detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]

93. Zhang, H.; Li, J.; Song, R.; Li, Y. Multi-Scale Structure-Conditioned Feature Transform Network for Object Detection in Remote Sensing Imagery. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 4208–4211.
94. Dong, Z.; Wang, M.; Wang, Y.; Zhu, Y.; Zhang, Z. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2104–2114. [CrossRef]
95. Meng, Y.B.; Wang, F.; Liu, G.H. Remote sensing multi-scale object detection based on multivariate feature extraction and characterization optimization. *Opt. Precis. Eng.* **2023**, *31*, 2465–2482. [CrossRef]
96. Yao, Q.L.; Hu, X.; Lei, H. Object Detection in Remote Sensing Images Using Multiscale Convolutional Neural Networks. *Acta Opt. Sin.* **2019**, *48*, 1266–1274.
97. Zhang, Y.Z.; Guo, W.; Li, W.B. Omnidirectional accurate detection algorithm for dense small objects in remote sensing images. *J. Jilin Univ.* **2023**, 1–9. [CrossRef]
98. Zhang, M.; Zheng, H.; Gong, M.; Wu, Y.; Li, H.; Jiang, X. Self-structured pyramid network with parallel spatial-channel attention for change detection in VHR remote sensed imagery. *Pattern Recognit.* **2023**, *138*, 109354. [CrossRef]
99. Zhou, J.; Zhang, R.; Zhao, W.; Shen, S.; Wang, N. APS-Net: An Adaptive Point Set Network for Optical Remote-Sensing Object Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 1–5. [CrossRef]
100. Wang, C.; Sun, W.; Fan, D.; Liu, X.; Zhang, Z. Adaptive feature weighted fusion nested U-Net with discrete wavelet transform for change detection of high-resolution remote sensing images. *Remote Sens.* **2021**, *13*, 4971. [CrossRef]
101. Xu, R.; Tao, Y.; Lu, Z.; Zhong, Y. Attention-mechanism-containing neural networks for high-resolution remote sensing image classification. *Remote Sens.* **2018**, *10*, 1602. [CrossRef]
102. Liu, M.; Jiao, L.; Liu, X.; Li, L.; Liu, F.; Yang, S. C-CNN: Contourlet convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 2636–2649. [CrossRef]
103. Hu, X.D.; Wang, X.Q.; Meng, F.J.; Hua, X.; Yan, Y.J.; Li, Y.Y.; Huang, J.; Jiang, X.L. Gabor-CNN for object detection based on small samples. *Def. Technol.* **2020**, *16*, 1116–1129. [CrossRef]
104. Chen, Y.; Zhu, L.; Ghamisi, P.; Jia, X.; Li, G.; Tang, L. Hyperspectral images classification with Gabor filtering and convolutional neural network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2355–2359. [CrossRef]
105. Zheng, S.; Wu, Z.; Xu, Y.; Wei, Z.; Plaza, A. Learning orientation information from frequency-domain for oriented object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [CrossRef]
106. El-Khamy, S.E.; Al-Kabbany, A.; Shimaa, E.L.B. MLRS-CNN-DWTPL: A new enhanced multi-label remote sensing scene classification using deep neural networks with wavelet pooling layers. In Proceedings of the 2021 International Telecommunications Conference (ITC-Egypt), Alexandria, Egypt, 13–15 July 2021; IEEE: Piscataway Township, NJ, USA, 2021; pp. 1–5.
107. El-Gayar, M.M. Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*. [CrossRef]
108. He, Y.; Zhou, S.; Quan, X. Remote Sensing Image Scene Classification Based on ECA Attention Mechanism Convolutional Neural Network. In Proceedings of the 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 12–14 October 2022; IEEE: Piscataway Township, NJ, USA, 2022; pp. 1265–1269.
109. Tsourounis, D.; Kastaniotis, D.; Theoharatos, C.; Kazantzidis, A.; Economou, G. SIFT-CNN: When Convolutional Neural Networks Meet Dense SIFT Descriptors for Image and Sequence Classification. *J. Imaging* **2022**, *8*, 256. [CrossRef] [PubMed]
110. Li, J.; Wang, T.; Gao, M.; Zhu, A.; Shan, G.; Snoussi, H. Two Stream Neural Networks with Traditional CNN and Gabor CNN for Object Classification. In Proceedings of the 2018 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 9350–9355.
111. Huo, C.; Zhou, Z.; Lu, H.; Pan, C.; Chen, K. Fast object-level change detection for VHR images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *7*, 118–122. [CrossRef]
112. Mboga, N.; Georganos, S.; Grippa, T.; Lennert, M.; Vanhuysse, S.; Wolff, E. Fully convolutional networks and geographic object-based image analysis for the classification of VHR imagery. *Remote Sens.* **2019**, *11*, 597. [CrossRef]
113. Zhang, Y.; Zhou, P.; Ren, Y.; Zou, Z. GPU-accelerated large-size VHR images registration via coarse-to-fine matching. *Comput. Geosci.* **2014**, *66*, 54–65. [CrossRef]
114. Saha, S.; Bovolo, F.; Bruzzone, L. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3677–3693. [CrossRef]
115. Chen, X.; Zhang, Q.; Han, J.; Han, X.; Liu, Y.; Fang, Y. Research progress of deep learning-based object detection of optical remote sensing image. *J. Commun.* **2022**, *43*, 190–203.
116. Heitz, G.; Koller, D. Learning spatial context: Using stuff to find things. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 30–43.
117. Tanner, F.; Colder, B.; Pullen, C.; Heagy, D.; Eppolito, M.; Carlan, V.; Oertel, C.; Sallee, P. Overhead imagery research data set—An annotated data library & tools to aid in the development of computer vision algorithms. In Proceedings of the 2009 IEEE Applied Imagery Pattern Recognition Workshop (AIPR 2009), Washington, DC, USA, 14–16 October 2009; IEEE Press: Piscataway Township, NJ, USA, 2009; pp. 1–8.
118. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium, Worth, TX, USA, 23–28 July 2017; IEEE Press: Piscataway Township, NJ, USA, 2017; pp. 3226–3229.

119. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; IEEE Press: Piscataway Township, NJ, USA, 2015; pp. 3735–3739.
120. Cheng, G.; Han, J.; Zhou, P.; Guo, L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS J. Photogramm. Remote Sens.* **2014**, *98*, 119–132. [CrossRef]
121. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small object detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [CrossRef]
122. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. In Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods, Porto, Portugal, 24–26 February 2017; SciTePress: Francisco, Italy, 2017; pp. 324–331.
123. Liu, K.; Mattyus, G. Fast multiclass vehicle detection on aerial images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1938–1942.
124. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [CrossRef]
125. Zhang, Y.; Yuan, Y.; Feng, Y.; Lu, X. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5535–5548. [CrossRef]
126. Zou, Z.X.; Shi, Z.W. Random access memories: A new paradigm for object detection in high resolution aerial remote sensing images. *IEEE Trans. Image Process.* **2018**, *27*, 1100–1111. [CrossRef] [PubMed]
127. Yang, Y.M. ITCVD Dataset[EB]. 2018. Available online: https://research.utwente.nl/en/datasets/itcvd-dataset (accessed on 28 November 2023).
128. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
129. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; IEEE Press: Piscataway Township, NJ, USA, 2018; pp. 3974–3983.
130. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 116–130. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.