The International Journal of Interdisciplinary Organizational Studies
ISSN: 2324-7649 (Print), ISSN: 2324-7657 (Online)
Volume 19, Issue 1, January-June, 2024

TRANSFER LEARNING IN CONJUNCTION WITH MULTI-OBJECT DETECTION USING YOLO RCNN

Kadapala Anjaiah
Research scholar in computer science and engineering, Osmania University, Hyderabad Telangana, Email: kadapala.anjaiah@gmail.com

Dr. K. Sagar.
Professor and vice principal, Sreyas institute of engineering and technology,
Mail: Sagar.k@sreyas.ac.in

Abstract: Processing images from satellite photography is a significant difficulty. Finding things in a satellite picture is a crucial job in this field. Since there has been a lot of study done on machine learning-based image processing, machine learning techniques may be used for this. Image processing tasks may be carried out via a multitude of machine learning-based supervised and unsupervised techniques. This research evaluated object identification systems based on machine learning using satellite photos. Finding one supervised and one unsupervised method to apply and compare across a created dataset was the aim of the study. In order to tackle these obstacles, our research used transfer learning in conjunction with multi-object identification deep learning algorithms using remotely sensed satellite data obtained on a diverse terrain. The models in the research were assessed using a fresh dataset of varied characteristics with five item classes that were gathered from Google Earth Engine at different places in the southern South African province of KwaZulu-Natal. The items in the dataset photos varied in terms of size and resolution. Using our recently developed dataset, five object identification techniques based on YOLO R-CNN architectures were examined via tests. A survey of the literature was done to determine the best algorithms for object recognition. For supervised learning, support vector machines as well as k-means were used, respectively, based on the literature review. To put these algorithms into practice, an experiment was conducted. A dataset of items from satellite pictures was produced specifically for the project. Both silhouette score analysis and confusion matrix analysis were used to assess the experiment's outcomes.The findings indicated that YOLOv8 had the quickest detection speed of 0.2 ms and the greatest detection accuracy of more than 90% for situations including vegetation and swimming pools. The study's findings indicate that, when it comes to object detection on satellite photos, support vector machines are more useful than k-means clustering.
Keywords: satellite images, object recognition, k-means clustering, support vector machines.

1 INTRODUCTION

One of the main areas of machine learning is image processing. In this arena, object recognition is a major difficulty [1]. Video surveillance, object tracking, and automated navigation are some of the fields in which object recognition is used [2]. When complicated pictures are taken into account, this becomes more difficult [2]. Experimental research has shown that complex natural pictures, such as satellite photographs, are beyond the capabilities of the traditional image processing methods [3]. Simultaneously, a wealth of satellite images is available [4]. Furthermore, weather, time, or environmental constraints do not impede the capture of distant sensor pictures [5]. This suggests that several distant sensor pictures can be generated on demand. Therefore, it is crucial to comprehend how effectively machine learning methods may be applied to them to conduct object detection.

Large amounts of data may be found in satellite photos. A number of applications, including natural resource management, spatial planning, and environmental monitoring, have prompted the development of techniques for extracting pictures from distant sensors [6]. Additionally, there is a tendency toward a growth in the quantity and complexity of remote sensor image-based urban applications such as vehicle detection and security [4], [6].

The need to handle huge data in a constrained amount of time is satisfied by object-based image analysis [6]. The analysis carried out on satellite pictures to perceive urban characteristics is reinforced by object-based categorization [7]. Applications for this are also found in the fields of metrology, agriculture, land use, and environmental monitoring [8].

Machine learning techniques may be used since there are many photos accessible. When machine learning methods are used, object detection in photos gets a high recognition rate [9]. This is particularly true for pictures that include intricate natural settings [10]. This categorization of complex natural pictures includes satellite photographs.

Applications ranging from mapping land cover for environmental monitoring, disaster management, and urban planning depend on remotely sensed satellite image processing [1]. To be more precise, the recognition of features like residential structures and bodies of water from remotely sensed satellite pictures is essential for managing and planning landscapes as well as averting and lessening natural calamities like fires and floods [2]. Nevertheless, owing to several difficulties such as enormous picture sizes, uneven illumination, and complicated backdrops, object recognition from satellite photos that are remotely sensed has proven to be challenging [3,4]. They have not proven useful for image analysis because there are not enough training datasets that accurately represent the complexity of landscapes [5]. Various CNN-based designs, including You Only Look Once (YOLO), Faster Region CNN (R-CNN), and Retina Net, have been suggested by researchers recently [6,7] for object recognition in satellite pictures. To maximize object recognition in photos, these designs use a variety of strategies, including feature pyramid networks, anchor boxes, and region proposal [7]. The well-liked Region-based Convolutional Neural Network (R-CNN) does a categorization after initially suggesting object areas [7, 8]. The adoption of deep learning techniques has been hindered by the limited training datasets and often complicated picture attributes, even though these approaches have shown considerable promise in order to recognize and locate visual things[9]. In order to overcome the difficulties associated with object identification and recognition in images and to create deeper learning models that are more reliable and accurate for pertinent applications, further research is thus required. Many challenges often accompany the use of deep learning algorithms for the identification and recognition of visual objects. Among these challenges include high item visual variability, noise including artifacts, as a lack of labeled data. [9, 10].
The methods used in this investigation are summed up as follows.
1. Creating a novel dataset by designing one using environmental perception data that is captured in various scenes and characterized by a variety of features;
2. Reviewing relevant literature;
3. Modeling R-CNN as well as YOLO-based algorithms using the new dataset;
4. Carrying out tests to ascertain how well the most sophisticated object identification algorithms perform in terms of object detection.

2 RELATED WORK

The precise identification of objects in remotely sensed pictures is essential for mapping and monitoring socioeconomic and biophysical aspects [18-21]. A technique for identifying buildings from images was presented by Wang et al. [22] and included combining a CNN with an LSTM network. The pictures' features were extracted using CNN, and the spatial connections between the features were modeled by LSTM. Furthermore, a strategy based on deep learning was put out to identify ships from synthetic aperture radar (SAR) pictures [23]. The suggested technique beat various deep learning systems that were already in use, according to the findings, which were obtained by using a region proposal network (RPN) and CNN to extract features from the photos[24-28]. In a factory, the model was used to identify objects in real time. An Auto-T-YOLO version of YOLOv4 was suggested by Sun et al. [29] to recognize objects in pictures.

Table 1 : lists the constraints and performance of object identification techniques on different datasets.

Methods	Datasets	Results Obtained	Limitations
FasterR-CNN [20]	AGs-GF1 & 2	86.0%, 12 fps	Lower classification accuracy
YOLOv3 [20]	AGs-GF1 7 2	90.4%, 73 fps	Slower classification rate
YOLOv4 [19]	MS-COCO	44.5%, 64.7 fps	Requires larger computational power
YOLOv5s [21]	SIMD	5.9 ms, 61.8 mAP	Lower recognition speed as well as precision

3 METHODOLOGY

An outline of the suggested technique is provided in this section (Figure 1). The process pipeline is shown in the image, which includes the model architectures, training and learning procedures for the models, and dataset generation and pre-processing. The classification output receives the anticipated end findings. This section provides further details on the model architectures utilized in this investigation. Five R-CNN and YOLO-based deep learning techniques have been tested in this work. The steps in our technique are outlined and spoken about below:

- Preparing and creating datasets.
- Model architectures.
- Training models using transfer learning.

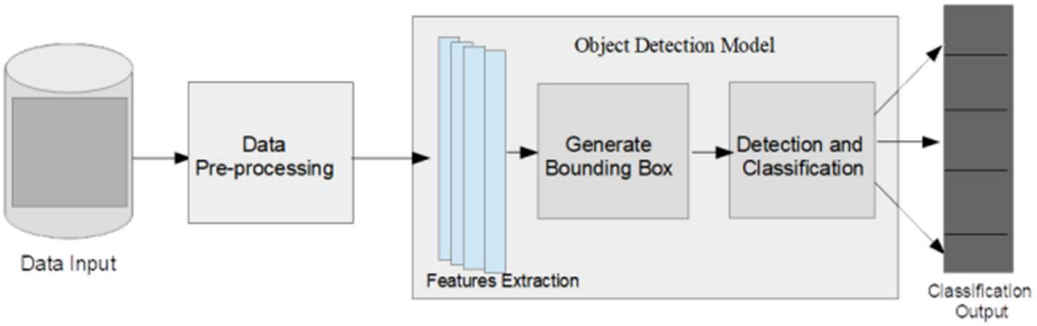

Figure 1. Architectural diagram for the proposed methodology.

3.1 Dataset Creation

The initial stage of the target detection process is the dataset development, illustrated in Figure 1. Deep learning model performance and accuracy are determined by the quantity and quality of the datasets used, which makes this procedure crucial. Extensive and high-quality datasets are necessary for object recognition in photos. The development is made using large-scale datasets with deep learning algorithms that recognize objects in precise places from photos has been spurred by recent technological advancements.A vision platform called Roboflow [19-25] was used to arrange, label, and compile the photos into datasets.

Ninety-two satellite photos that were marked up with the use of the multi-class categorization methodology make up the dataset. Five items may be seen in the images: a house, roads, a beach, a swimming pool, and flora. Figure 4 shows some sample photos from the collection. The dataset is divided as training, testing, validation. 61 photos along with one annotation file made up the training set; 21 photos and single annotation file made up the validation set; and 10 pictures made up the testing set. The photos underwent a few preparation operations, such as pixel data auto-orientation. To improve the amount, picture augmentation was used to the training dataset. With the suggested dataset, we used a data augmentation technique to improve the models' resilience and accuracy of detection. The method uses geometric modifications and augmentations to alter the spatial orientation of pictures without altering their content. To generate a mirror image, flip the picture horizontally. To invert the image, flip it vertically. Additionally, in order to replicate the various viewing angles of an item in the picture, the images are rotated by 90, 180, or 270 degrees. These procedures were run through 100 times in order to double the number of training photos.

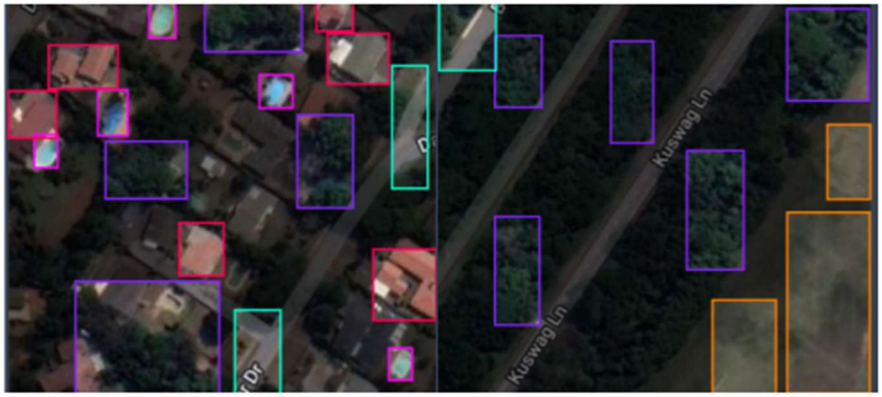

Figure 2: Labeling images for objects detection.

The model series utilised in the research are listed in Table 2 and include YOLO versions v5–v8 and Detectron2, an R-CNN-based model. Pascal VOC and the Common Object in Context (COCO) datasets are among the pretrained datasets used for Detectron2 architecture, which is often built on R-CNN frameworks [21]. The Detectron2 system consists of the detecting head, data loading, feature pyramid network (FPN), region proposal network (RPN) and backbone network.

The head, neck, and backbone are the three basic elements that make up the YOLO series. The head produces predictions for object recognition, the neck integrates information from various backbone levels, and the input picture's characteristics are extracted by the backbone. At the backbone segment, YOLOv7 uses an efficient layer aggregation network and models based on concatenation. In the neck area, it also makes use of Path Aggregation Network (PAN)-based FPN and has two heads: lead and auxiliary heads. Anchor-free object recognition, which does away with the necessity mechanism, are used by YOLOv5 and YOLOv8.

By substituting a unique modules in the neck and backbone regions, the YOLOv8 design offers a framework with fewer parameters and tensors overall. As shown by our findings, this has resulted in YOLOv8 detection that is effective, quicker calculation, and a reduction in computing resources.

To increase our models' efficacy, we used a transfer learning strategy in this work [29]. To be more precise, this meant pre-training the models using a bigger dataset that is made up of many pictures of everyday things. We might make use of the abundance of data for training with our mathematical models on a big and diverse collection of data in order to produce a feature extractor that is more broadly applicable. We pre-trained the models and then used our freshly generated dataset to retrain them. In this project, the main target tasks for customisation were the classes in the suggested dataset. We retrained the models on these specific classes in order to improve their precision and suitability for the targeted tasks. All in all, this strategy allowed us to make the most of the enormous quantity of data at our disposal and produce a powerful feature extractor. This thus made it possible for us to more effectively fine-tune our models to provide improved accuracy and generalization on the recently generated dataset.

4 RESULTS AND DISCUSSION

4.1 Evaluation Metrics

Our unique dataset was used to test the object identification techniques based on deep learning. Five measures were used to analyze the models: recall (R), mean average precision (mAP), average precision (AP), precision (P), and detection accuracy (DA). These are shown as follows:

The degree to which a model can accurately identify items in a picture is measured by its detection accuracy. A model's detection accuracy is assessed using a number of measures, such as precision, recall, and F1-score.

The percentage of true positives, or items that the model properly detects, among all the objects it detects is referred to as precision. A high precision means that the majority of the items the model identifies are accurate, indicating a low incidence of false positives;

The recall metric quantifies the percentage of real items in the picture that the model correctly identified as true positives. A high recall means that the majority of the items in the picture can be identified by the model.

4.2 Results Discussion

The precision, recall, mAP50, and mAP (50:95) measures were utilized to assess verification performance over the proposed dataset. The outcomes are shown in Table 2, where it is evident that the most advanced detection techniques outperform one another in terms of every assessment criteria. In particular, YOLOv8 maintained the maximum speed restriction of 0.2 ms and obtained 70%, 62%, 45%, and 19.5% in accuracy, mAP50, recall, and mAP (52:97), respectively. YOLOv5, YOLOv6, and YOLOv7 behaved similarly, although showing a little improvement. A 50% accuracy score was attained by Detectron2, but more slowly than by the YOLO-based models.

Table 2. Performance of the models on the suggested dataset

Methods	Precision	Recall	mAP50	MAP50-95
Detectron_2	50	33.7	17	25
YOLOv_5	54.2	49.2	28	19.4
YOLOv_6	54.2	48.3	22	19.9
YOLOv_7	55.3	47.2	24	21.0
YOLOv_5	70	62	19.5	17.5

A comparative examination of several models evaluated on two datasets, VisDrones [25] and Pascal VOC2007 [26], revealed that YOLOv8 performed better and had the greatest mAP value.

Comparative Study of the Algorithms on Openly Accessible Datasets: Pascalvoc and Visdrones

A comparative examination of several models evaluated on two datasets, VisDrones [29] and Pascal VOC2007 [30], revealed that YOLOv8 performed better and had the greatest mAP value. Table 3 provides an overview of these datasets, and the outcomes of testing YOLO_v3, YOLO_v5, YOLO_v7, and YOLO_v8 on the corresponding datasets. The results show that YOLOv8 performs better than the other two models, even when it comes to tiny and typical target items.

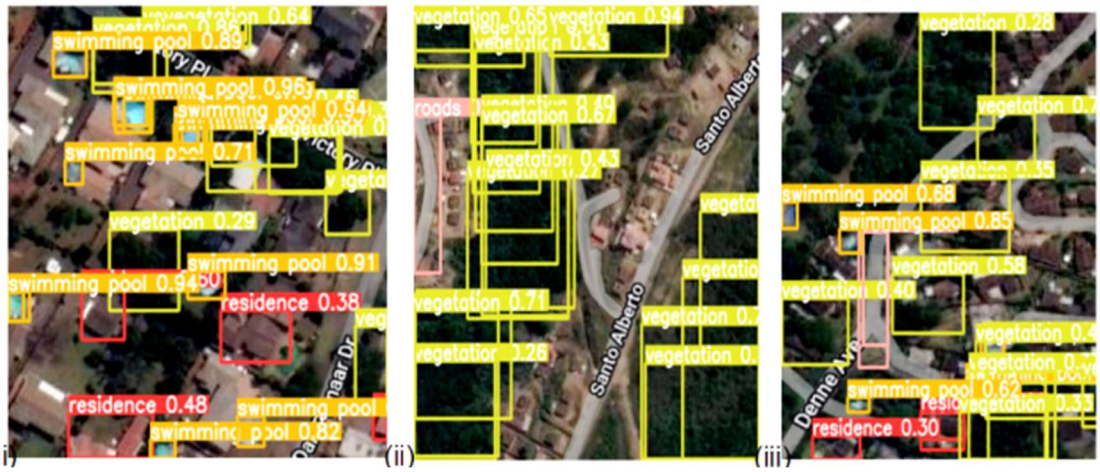

Figure 3. Finding dense items in test datasets of sample satellite images; (i) Finding residences, vegetation, and swimming pools; (ii) Finding objects, such vegetation; (iii) Finding vegetation and pools.

Figure 4. Detection of Densely Distributed Objects from Sample Satellite communication.

Table 3. YOLOv8's classwise performance on the suggested dataset.

Label	Precision	Recall	mAP50	MAP50-95
Residence	42.2	43.2	20.1	13.9
Roads	43.1	58.2	14.8	4.98
Shorelines	55.6	97.2	99.9	60.32
Swimming Pool	63.8	65.2	46.2	13.9
Vegetation	58.6	63.5	13.2	9.63

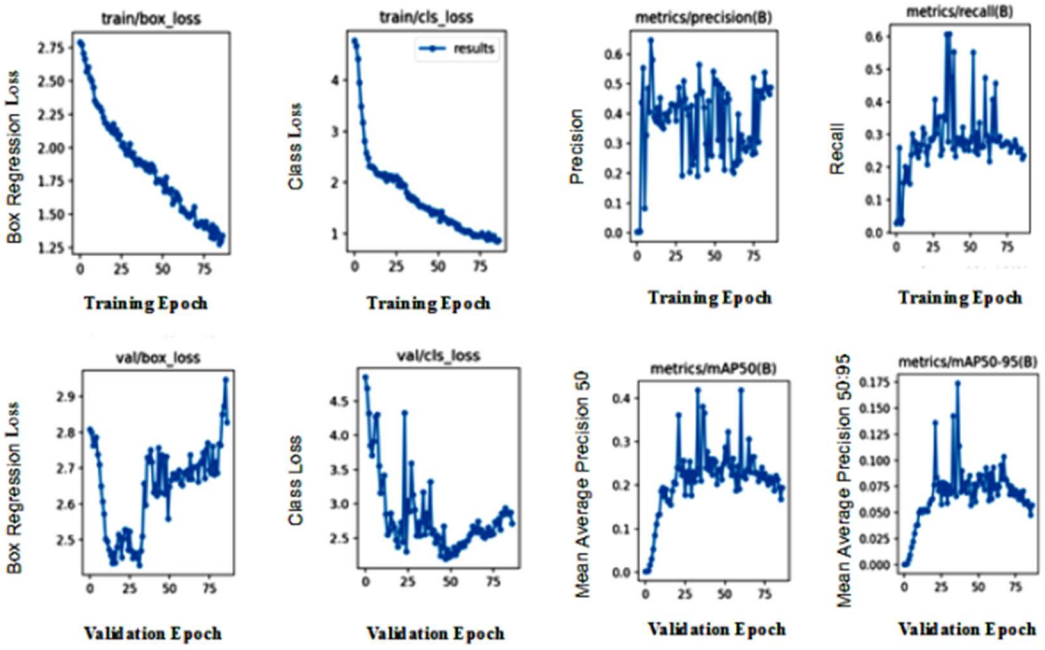

Figure 5. Performance assessment curves for YOLOv8 using the suggested dataset; Training curves for recall, accuracy, class loss, and box regression loss are shown in the first row; Validation curves for class loss, and accuracy.

5 CONCLUSIONS

The article discusses the findings from an investigation that was conducted to test the efficiency of cutting-edge technologies to detect objects using satellite remote-sensing images. The study revealed the existence of a variety of complicated aspects, like large intra-class variations, significant inter-class similarities, as well as an unforgiving background, hindered the present state of technology for object recognition on images from satellites that are aerial. This paper presents a brand new dataset based on experimental research using an approach to transfer learning that provides an extensive analysis and analysis of current deep learning based techniques for the detection of objects with high-resolution imagery. When tested on the suggested dataset, the performance of these approaches using YOLOv8 is encouraging, obtaining 68% and 60% in accuracy and recall. On the Pascal VOC and Visdrone datasets, a comparative study of object identification techniques was also conducted, with YOLOv8 demonstrating better performance.

Future Works

Despite the fact that this study has been able to address the issues raised, some things went unnoticed. Very tiny objects, especially those grouped together with certain occlusions, are still difficult for the suggested object identification framework to identify. Improving localization precision is a challenging endeavor. Future research will look at ways to improve detection precision and localization even further for really small objects when challenging occlusions are present.

Limitations:

Restricted labelled dataset: To improve the variety of training data, the dataset underwent data augmentation. The method modifies the spatial orientation of pictures by applying geometric augmentations and transformations. The procedures were run through 100 times in order to double the number of training photos. In order to improve performance, this research additionally used a transfer learning as well as dynamic data fusion methodology to model the object detection method on the recently obtained dataset. Using large-scale datasets and pre-trained models, the transfer learning technique enhances the effectiveness of object identification and recognition in imagery obtained from satellite pictures.

6 REFERENCES

1. Simelane, S.P.; Hansen, C.; Munghemezulu, C. The use of remote sensing and GIS for land use and land cover mapping in Eswatini: A Review. S. Afr. J. Geomat. 2022, 10, 181–206. [CrossRef]
2. Bhuyan, K.; Van Westen, C.; Wang, J.; Meena, S.R. Mapping and characterising buildings for flood exposure analysis using open-source data and artificial intelligence. Nat. Hazards 2022, 1–31. [CrossRef]
3. Qi, W. Object detection in high resolution optical image based on deep learning technique. Nat. Hazards Res. 2022, 2, 384–392. [CrossRef]
4. Vemuri, R.K.; Reddy, P.C.S.; Kumar, B.S.P.; Ravi, J.; Sharma, S.; Ponnusamy, S. Deep learning based remote sensing technique for environmental parameter retrieval and data fusion from physical models. Arab. J. Geosci. 2021, 14, 1230. [CrossRef]

5. 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]
6. Li, Z.; Wang, Y.; Zhang, N.; Zhang, Y.; Zhao, Z.; Xu, D.; Ben, G.; Gao, Y. Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey. Remote Sens. 2022, 14, 2385. [CrossRef]
7. Karim, S.; Zhang, Y.; Yin, S.; Bibi, I.; Brohi, A.A. A brief review and challenges of object detection in optical remote sensing imagery. Multiagent Grid Syst. 2020, 16, 227–243. [CrossRef]
8. Pham, M.-T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-Stage Detector of Small Objects under Various Backgrounds in Remote Sensing Images. Remote Sens. 2020, 12, 2501. [CrossRef]
9. Nawaz, S.A.; Li, J.; Bhatti, U.A.; Shoukat, M.U.; Ahmad, R.M. AI-based object detection latest trends in remote sensing, multimedia and agriculture applications. Front. Plant Sci. 2022, 13, 1041514. [CrossRef]
10. Liu, J.; Yang, D.; Hu, F. Multiscale object detection in remote sensing images combined with multi-receptive-field features and relation-connected attention. Remote Sens. 2022, 14, 427. [CrossRef]
11. Ahmed, M.;Hashmi, K.A.; Pagani, A.; Liwicki, M.; Stricker, D.; Afzal, M.Z. Survey and performance analysis of deep learning based object detection in challenging environments. Sensors 2021, 21, 5116. [CrossRef]
12. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Proceedings, Part V 13, Zurich, Switzerland, 6–12 September 2014; Springer International Publishing: New York, NY, USA, 2014; pp. 740–755;
13. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. 2010, 88, 303–338. [CrossRef]
14. Zhang, X.; Han, L.; Han, L.; Zhu, L. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sens. 2020, 12, 417. [CrossRef]
15. Wang, Y.; Xu, C.; Liu, C.; Li, Z. Context Information Refinement for Few-Shot Object Detection in Remote Sensing Images. Remote Sens. 2022, 14, 3255. [CrossRef]
16. Chang, R.-I.; Ting, C.; Wu, S.; Yin, P. Context-Dependent Object Proposal and Recognition. Symmetry 2020, 12, 1619. [CrossRef]
17. Saito, S.; Yamashita, T.; Aoki, Y. Multiple object extraction from aerial imagery with convolutional neural networks. Electron.Imaging 2016, 2016, 010402-1–010402-9.
18. Xiaolin, F.; Fan, H.; Ming, Y.; Tongxin, Z.; Ran, B.; Zenghui, Z.; Zhiyuan, G. Small object detection in remote sensing images based on super-resolution. Pattern Recognit. Lett. 2022, 153, 107–112. [CrossRef]
19. Tong, K.; Wu, Y.; Zhou, F. Recent advances in small object detection based on deep learning: A review. Image Vis. Comput. 2020, 2, 103910. [CrossRef]
20. Li, M.; Zhang, Z.; Lei, L.; Wang, X.; Guo, X. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. Sensors 2020, 2, 4938. [CrossRef]
21. Wan, D.; Lu, R.; Wang, S.; Shen, S.; Xu, T.; Lang, X. YOLO-HR: Improved YOLOv5 for Object Detection in High-Resolution Optical Remote Sensing Images. Remote Sens. 2023, 2, 614. [CrossRef]

22. Wang, Y.; Gu, L.; Li, X.; Ren, R. Building extraction in multitemporal high-resolution remote sensing imagery using a multifeature LSTM network. IEEE Geosci. Remote Sens. Lett. 2020, 2, 1645–1649. [CrossRef]
23. Gao, F.; He, Y.; Wang, J.; Hussain, A.; Zhou, H. Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images. Remote Sens. 2020, 2, 2619. [CrossRef]
24. Gan, Y.; You, S.; Luo, Z.; Liu, K.; Zhang, T.; Du, L. Object detection in remote sensing images with mask R-CNN. In Journal of Physics: Conference Series; IOP Publishing: Bristol, UK, 2020; Volume 1673, p. 012040.
25. 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, 2, 2486–2498. [CrossRef]
26. Van Etten, A. Satellite imagery multiscale rapid detection with windowed networks. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 7–11 January 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 735–743.
27. Van Etten, A. You only look twice: Rapid multi-scale object detection in satellite imagery. arXiv 2018, arXiv:1805.09512.
28. Ku, B.; Kim, K.; Jeong, J. Real-Time ISR-YOLOv4 Based Small Object Detection for Safe Shop Floor in Smart Factories. Electronics 2022, 2, 2348. [CrossRef]
29. Sun, B.; Wang, X.; Oad, A.; Pervez, A.; Dong, F. Automatic Ship Object Detection Model Based on YOLOv4 with Transformer Mechanism in Remote Sensing Images. Appl. Sci. 2023, 2, 2488. [CrossRef]
30. Yu, L.; Wu, H.; Liu, L.; Hu, H.; Deng, Q. TWC-AWT-Net: A transformer-based method for detecting ships in noisy SAR images. Remote Sens. Lett. 2023, 2, 512–521. [CrossRef]