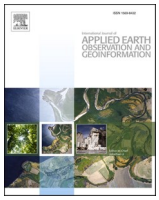
International Journal of Applied Earth Observation and Geoinformation 134 (2024) 104208

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Object detection in aerial images using DOTA dataset: A survey

Ziyi Chen [a], Huayou Wang [a], Xinyuan Wu [b], Jing Wang [a], Xinrui Lin [c], Cheng Wang [d], Kyle Gao [e], Michael Chapman [f], Dilong Li [a,*]

[a] Department of Computer Science and Technology, Fujian Key Laboratory of Big Data Intelligence and Security, Xiamen Key Laboratory of Data Security and Blockchain Technology, Huaqiao University, 668 Jimei Road, Xiamen, FJ 361021, China
[b] Institute of Advanced Technology, University of Science and Technology of China, Hefei, AH 230088, China
[c] University of Edinburgh, Edinburgh EH8 9YL, United Kingdom
[d] School of Informatics, Xiamen University, Xiamen, FJ 361005, China
[e] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[f] Department of Civil Engineering Toronto Metropolitan University 350 Victoria St, Toronto, Ontario M5B 2K3, Canada

ARTICLE INFO

Keywords:
Object detection
RSIs
Deep learning
DOTA dataset

ABSTRACT

In recent years, the Dataset for Object deTection in Aerial images (DOTA) dataset has played a pivotal role in advancing object detection in aerial images (ODAI). Despite its significance, there hasn't been a comprehensive review summarizing its research developments. Addressing this gap, this paper offers the first comprehensive overview on the subject. Within this review, we begin by examining prevalent object detection datasets of natural scene images alongside object detection datasets of remote sensing images (RSIs). We then present an in-depth comparative analysis between these datasets and the DOTA dataset, supported by numerous charts and tables. We proceed to outline both traditional techniques for ODAI and methods rooted in deep learning. Subsequently, we provide a recap of the latest advancements in the field achieved using the DOTA dataset. Concluding our review, we delve into the current challenges facing ODAI and propose potential future research directions.

1. Introduction

Owing to the distinctive characteristics of remote sensing images (RSIs), object detection in aerial images (ODAI) holds significant importance across various sectors, including intelligent monitoring, urban planning, and precision agriculture (Li et al. 2019b). As remote sensing technology advances rapidly, we can obtain aerial and satellite images, enabling widespread observation and monitoring of the Earth's surface. Currently, deep learning models, with their strong feature learning capabilities, have achieved significant strides in numerous fields, including ODAI (Yang et al., 2021e; Yang et al. 2021d; Sun et al. 2018; Yang et al., 2023b; Xie et al. 2021). However, the efficacy of these models largely hinges on ample training data, underscoring the paramount importance of datasets in this context.

A high-quality dataset should exhibit diversity and comprehensiveness to capture a wide range of real-world scenarios and target categories. Currently, while there are popular remote sensing image datasets proposed, such as SZTAKI-INRIA (Benedek et al. 2011), Vehicle Detection in Aerial Imagery (VEDAI) (Razakarivony and Jurie 2016), High-resolution ship collections 2016 (HRSC2016) (Liu et al. 2016), and Fine-Grained Ship Detection (FGSD) (Chen et al. 2020a), they possess inherent limitations such as restricted sample sizes and lack of comprehensive annotations. These shortcomings hinder their capability to fulfill the data prerequisites for remote sensing image-based model training comprehensively. To address these dataset challenges, Xia et al. (2018) presented the Dataset for Object deTection in Aerial images (DOTA) dataset. This dataset is characterized by several salient features. It comprises an extensive collection of high-resolution RSIs that encompass diverse target categories, including vehicle, bridge, ship, etc. Furthermore, the DOTA dataset's objects manifest a plethora of shapes, scales, and orientations, posing challenges to the robustness of detection methodologies. Moreover, DOTA ensures detailed object annotations, inclusive of bounding box coordinates, category labels, and orientation angles, thereby establishing a robust benchmark for algorithmic

* Corresponding author.
E-mail addresses: chenziyihq@hqu.edu.cn (Z. Chen), 22014083052@stu.hqu.edu.cn (H. Wang), xinyuanwu02@163.com (X. Wu), wroaring@hqu.edu.cn (J. Wang), X.Lin-35@sms.ed.ac.uk (X. Lin), cwang@xmu.edu.cn (C. Wang), y56gao@uwaterloo.ca, y56gao@uwaterloo.ca (K. Gao), mchapman@torontomu.ca (M. Chapman), scholar.dll@hqu.edu.cn (D. Li).

https://doi.org/10.1016/j.jag.2024.104208
Received 29 July 2024; Received in revised form 23 September 2024; Accepted 4 October 2024
Available online 20 October 2024

evaluation and comparative analysis.

The DOTA dataset has garnered considerable attention from researchers, leading to a series of significant research outcomes in the field of ODAI. To delve deeper into the developmental trends of this domain, we employed Web of Science and Google Scholar as search engines, retrieving over 1700 journal papers related to the DOTA dataset. However, based on our survey, there has yet to be a comprehensive literature review summarizing the research progress of the DOTA dataset. Given this context, this paper aims to offer an in-depth review of existing studies, highlight breakthrough advancements made on the DOTA dataset in recent years, and analyze the challenges currently faced by the ODAI domain. Building upon an analysis of paper titles, abstracts, and the like, this paper has selectively reviewed approximately 110 research papers closely related to the DOTA dataset published from 2018 to June 2024, which forms the crux of this review. Fig. 1 presents the annual publication statistics regarding the DOTA dataset since 2018.

Our major contributions are as follows:

First, we present a comprehensive overview of literature studies related to the DOTA dataset for the first time. We have conducted an exhaustive review of research conducted on the DOTA dataset over the past six years and summarized the latest advancements achieved in the realm of ODAI using this dataset.

Second, this paper goes beyond the DOTA dataset by revisiting other prevalent datasets related to natural scene images and RSIs, and offers a detailed comparative analysis against the DOTA dataset.

Third, this paper elaborates on both traditional object detection techniques and those rooted in deep learning specific to remote sensing imagery. We delve into the various challenges faced in ODAI and discuss methodologies to address them.

The rest of this paper is structured as follows. Section 2 provides an overview of object detection datasets in both natural scene images and RSIs, with a specific emphasis on the DOTA dataset. Section 3 reviews traditional object detection techniques in RSIs alongside those leveraging deep learning. Section 4 encapsulates the latest research milestones achieved using the DOTA dataset in the domain of ODAI, followed by a synthesis and brief on the carefully selected more than 110 articles. Section 5 delves deeper into the challenges currently pervasive in the field of ODAI and propose potential future research directions. Finally, Section 6 concludes the paper. The organization of this survey is presented in Fig. 2.

2. Object detection datasets

With the advent of deep learning techniques, various datasets have laid a solid foundation for the research in object detection algorithms. Not only have these datasets established benchmarks for algorithmic assessment, but they have also catalyzed the emergence of novel methodologies and technologies. This section first reviews the object detection datasets from natural scene images and details comparative analysis with the DOTA dataset, elucidating the differences between RSIs and natural scene images. Subsequently, we review several prevalent remote sensing image datasets and compare them with the DOTA dataset, highlighting the unique advantages of the DOTA dataset. Finally, the DOTA dataset is introduced.

2.1. Object detection datasets of natural scene images

Numerous natural scene image datasets are available in the literature for object detection, widely utilized in this task. These datasets typically comprise RGB bands and do not include geo-coordinates. These datasets include PASCAL Visual Object Classes (VOC) (Everingham et al. 2010), Microsoft Common Object in COntext (MS COCO) (Lin et al. 2014), Large Vocabulary Instance Segmentation (LVIS) dataset (Gupta et al. 2019) and ImageNet (Deng et al. 2009), which are most widely used.

PASCAL VOC is a classic computer vision dataset series, dedicated to advancing research and development in tasks such as object detection and image segmentation. Specifically, the PASCAL VOC 2007 dataset (Everingham et al. 2010) includes 20 different object categories, encompassing common objects like people, vehicles, and animals. This dataset contains 9,963 images, with 5,011 images designated for training and 4,952 images for testing. The PASCAL VOC 2012 dataset (Everingham et al. 2015) is an expanded version of the PASCAL VOC series, retaining the 20 categories. It comprises 11,540 images for training and 23,080 images for testing.

MS COCO is a larger-scale multi-task computer vision dataset compared to PASCAL VOC, aiming to foster research in object detection, image segmentation, and human pose estimation. The MS COCO dataset encompasses 80 distinct object categories, including person, animal, vehicle, etc. It consists of over 200,000 images, with approximately 80 K, 40 K, and 80 K allocated for training, validation, and testing respectively.

ImageNet is among the largest object detection datasets available, extensively utilized for tasks such as image classification, object localization, and object detection. The ImageNet dataset has had a profound impact in the field of computer vision through its image classification subset, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2014). The dataset covers 200 diverse object categories, containing over 500,000 images. Specifically, it

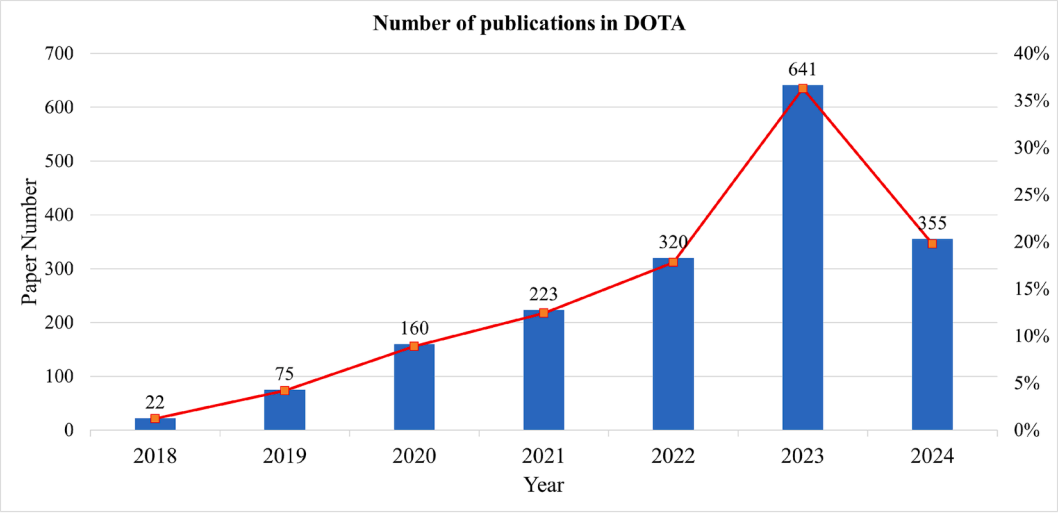


Fig. 1. The number of journal articles related to the DOTA dataset from 2018 to June 2024.

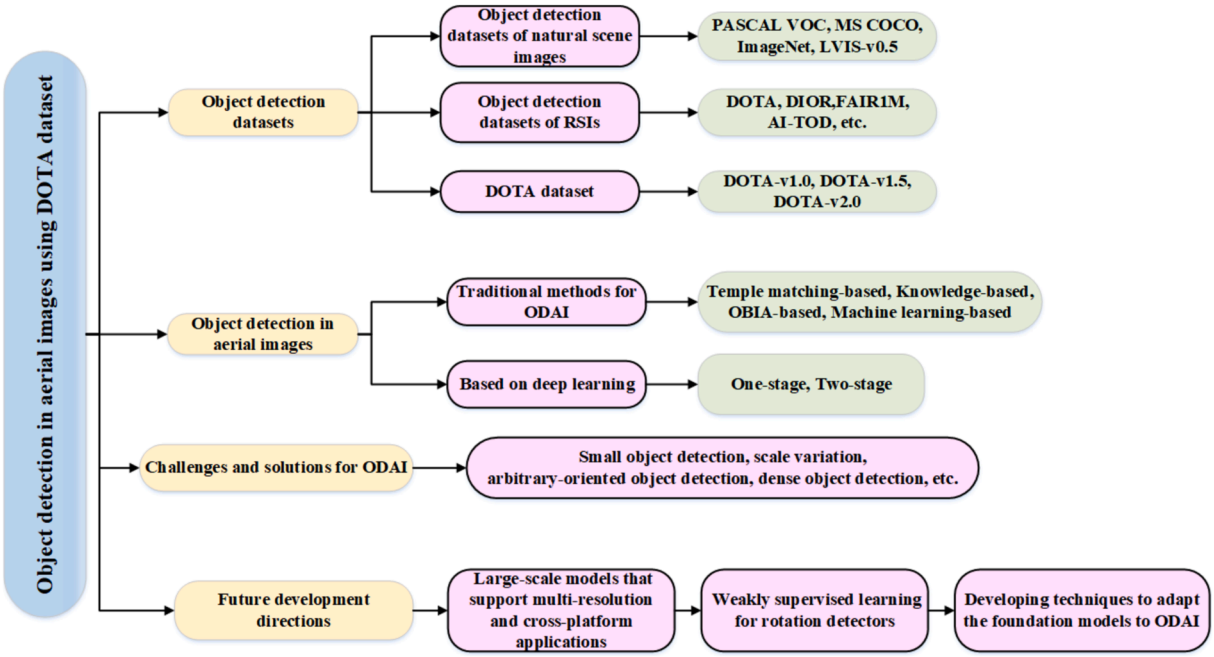


Fig. 2. The organization of this paper.

includes 456,567 images for training, 20,121 images for validation, and 40,152 images for testing.

LVIS is a recent benchmark designed for large vocabulary long-tail instance segmentation. It utilizes source images from the COCO dataset, with annotations created through an iterative object spotting method that naturally reflects the long-tail distribution of categories in images. The latest released version, v0.5, includes 1,230 object classes in the training set and 830 in the validation set, with the number of classes in the test set yet to be disclosed. These three sets comprise approximately 50,000, 5,000, and 20,000 images, respectively. We have summarized the statistical analysis results of other general object detection datasets and the DOTA dataset's training, validation, and testing sets in Table 1.

2.2. Object detection datasets of RSIs

In optical remote sensing, images frequently exhibit varying resolutions, leading to discrepancies in the amount of information they carry. Concurrently, terrains from diverse regions such as cities, deserts, and forests display distinct surface features, including but not limited to topography and vegetation cover. Hence, solely relying on the number of images to gauge the scale of a dataset seems inadequate.

In this context, to evaluate the scale of a dataset more objectively, we should also consider the pixel area of images and the variability of regional features. The pixel area of an image can intuitively reflect the amount of information it encompasses, especially in high-resolution images where the details of instances and contextual information become crucial for remote sensing tasks. Concurrently, the diversity in regional characteristics provides vital insights into the content's variety and complexity within the dataset. Thus, by considering these two dimensions collectively, we can gain a more comprehensive understanding of the dataset's informational content and further assess its potential value in practical applications (Xia et al. 2018).

Table 2
DOTA versus object detection datasets of natural scene images.

Datasets	Category	Image quantity	Bounding box quantity	Average bounding box quantity	Megapixel area
PASCAL	20	21,503	52,090	2.42	5,133
VOC	80	123,287	886,266	7.19	32,639
MS COCO	200	349,319	478,806	1.37	82,820
ImageNet					
DOTA-v1.0	15	2,806	188,282	67.10	19,173
DOTA-v1.5	16	2,806	402,089	143.73	19,173
DOTA-v2.0	18	11,268	**1,793,658**	**159.18**	**126,306**

Table 1
Comparison among DOTA and other general object detection datasets.

Datasets	Year	Train images	Validation images	Train and Val images	Test images	Total
VOC-2007	2007	2,501	2,510	5,011	4,952	9,963
VOC-2012	2012	5,717	5,823	1,540	10,991	22,531
ILSVRC-2014	2014	**456,567**	20,121	**476,668**	40,152	87,820
ILSVRC-2017	2017	**456,567**	20,121	**476,688**	65,500	113,168
MS COCO-2015	2015	82,783	**40,504**	123,287	**81,434**	**204,721**
MS COCO-2017	2017	118,287	5,000	123,287	40,670	163,957
LVIS-v0.5	2019	50,000	5,000	55,000	20,000	75,000
DOTA-v1.0	2018	1,411	458	1,869	937	2,806
DOTA-v1.5	2019	1,411	458	1,869	937	2,806
DOTA-v2.0	2021	1,830	593	2,423	8,845	11,268

The comparison results between the DOTA dataset and the object detection datasets from natural scene images are shown in Table 2. Some data in the table are referenced from the literature (Ding et al. 2021). From the table, we can observe that the DOTA dataset has a scale comparable to other large-scale datasets. However, in terms of the average number of instances per image, the DOTA dataset surpasses other object detection datasets from natural scene images. In the table, the PASCAL VOC dataset is marked as 07++12, indicating that it used the entire VOC2007 dataset and the Train and Validation sets of VOC2012 for training, while the Test sets of VOC2012 was used for testing since the VOC2012 Test set has not been made public. The MS COCO dataset data refers to the 2014 version of Train and Validation sets. The ImageNet dataset data is based on the 2017 version of the Train sets.

While significant success has been achieved in object detection within natural scene images, directly applying deep learning-based object detection methods to RSIs brings significant challenges. This stems from the fact that many datasets in the remote sensing domain are curated to cater to the specific needs of particular application scenarios (Xia et al. 2018). Fig. 3 displays sample images from the DOTA dataset.

In the Earth Observation field, several common remote sensing image datasets have been introduced, such as UCAS-AOD (Zhu et al. 2015), NWPU VHR-10 (Cheng et al. 2014), RSOD (Long et al. 2017), LEVIR (Zou and Shi 2017), and HRRSD (Zhang et al. 2019b). Although these datasets encompass multiple object categories, the number of samples they contain is relatively limited, making it challenging to effectively train robust deep models. Taking the RSOD dataset as an example, it consists of merely 976 images with 6,950 target instances, indicating that they still fall short of meeting the current requirements in the domain of ODAI.

To address the aforementioned issues, Xia et al. (2018) introduced the DOTA-v1.0 dataset in 2018. This dataset comprises 15 distinct object categories, encompassing a total of 188,282 instances across 2,806 images. Of these, 1,141 images are designated for training, 458 for validation, and 937 for testing. In the same year, Lam et al. (2018) presented the xView dataset, while Zhu et al. (2018) proposed the VisDrone dataset. In 2019, Li et al. (2019b) introduced the DIOR dataset. The xView dataset comprises 16 primary categories and 60 fine-grained categories, with a total of 1 million instances and 1,413 images. The VisDrone dataset encompasses 11 distinct object categories, with 540 k annotations, aggregating 54,200 instances and 10,209 images. Conversely, the DIOR dataset covers 20 different object categories, accounting for 190,288 instances and 23,463 images.

Although these datasets all fall within the domain of remote sensing imagery, the instance annotations in both the DIOR and xView datasets are based on Horizontal Bounding Boxes (HBB), which are not suitable for precise detection of objects in arbitrary orientations in RSIs. On the other hand, the VisDrone dataset primarily focuses on unmanned aerial vehicle (UAV) imagery and, while also being a large-scale dataset, places greater emphasis on the field of video object detection and tracking. Table 3 presents a comparison between the DOTA dataset and remote sensing image datasets, with some of the data in the table being referenced from the literature (Ding et al. 2021).

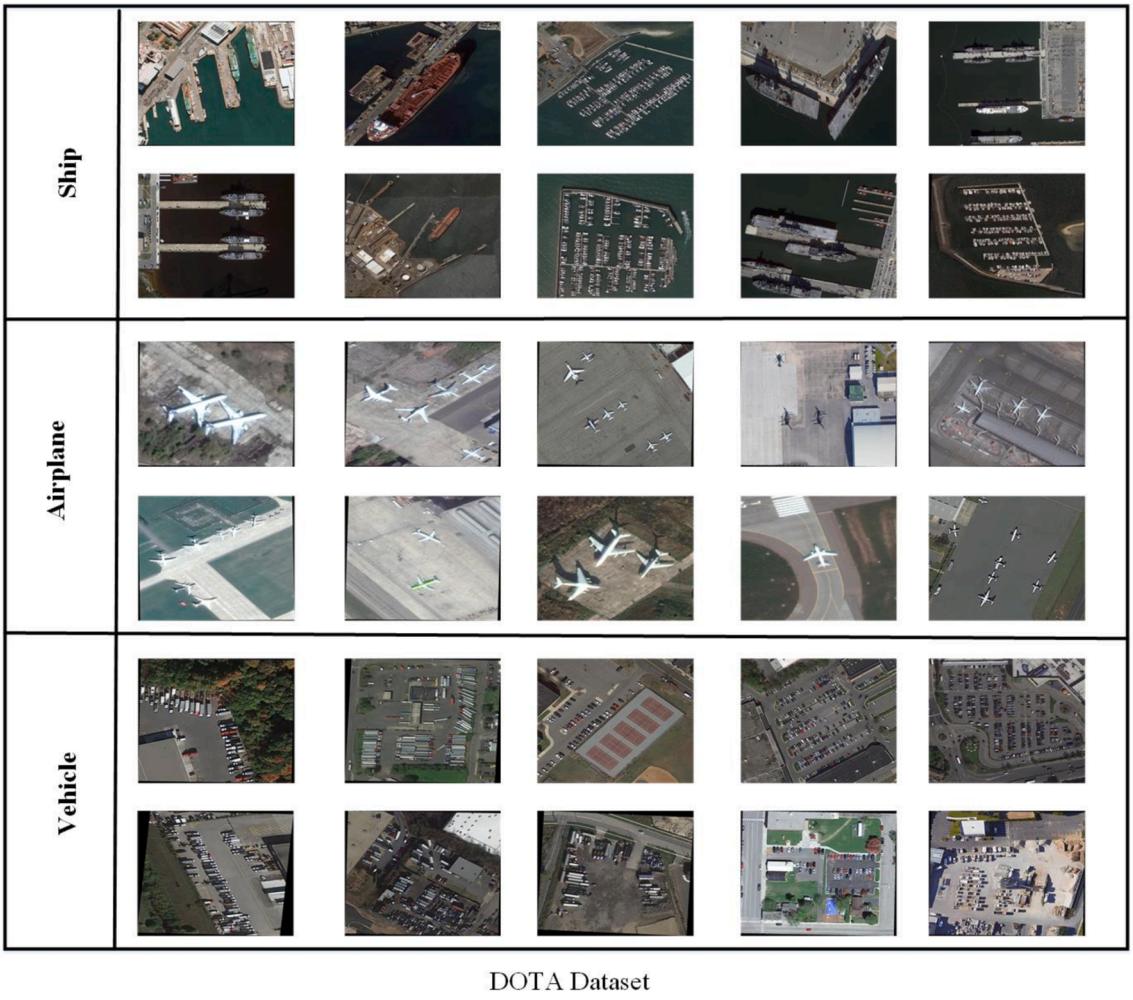


Fig. 3. Some examples taken from the DOTA dataset.

Table 3
DOTA versus object detection datasets in RSIs.

Datasets	Year	Publication	Annotation way	Image quantity	Categories	Instances	Image width
DOTA-v2.0 (Ding et al. 2021)	2021	TPAMI	oriented bounding box	11,268	18	**1,793,658**	800–20,000
DOTA-v1.5	2019	—	oriented bounding box	2,806	16	402,089	800–13,000
DOTA-v1.0 (Xia et al. 2018)	2018	CVPR	oriented bounding box	2,806	15	188,282	800–13,000
FAIR1M (Sun et al. 2022)	2022	ISPRS	oriented bounding box	42,796	37	1.02 million	600–10,000
AI-TOD (Wang et al. 2021)	2021	ICPR	horizontal bounding box	28,036	8	700,621	800
RarePlanes (Shermeyer et al. 2021)	2021	CVPR	Polygon	50,253	**110**	644,258	1,080
iSAID (Waqas Zamir et al. 2019)	2019	CVPR	Polygon	2,806	15	655,451	800–13,000
DIOR (Li et al. 2019b)	2019	ISPRS	horizontal bounding box	23,463	20	190,288	800
HRRSD (Zhang et al. 2019b)	2019	TGRS	horizontal bounding box	21,761	13	55,740	152–10,569
SpaceNet MVOI (Weir et al. 2019)	2019	ICCV	Polygon	60,000	1	126,747	900
LEVIR (Zou and Shi 2017)	2017	TIP	horizontal bounding box	22,000	3	11,000	600–800
UAVDT (Du et al. 2018)	2018	ECCV	horizontal bounding box	**78,320**	3	885,192	1,080
ITCVD (Yang et al. 2018)	2018	ICIP	horizontal bounding box	23,543	1	228	5,616
CARPPK (Hsieh et al. 2017)	2017	ICCV	horizontal bounding box	1,448	1	89,777	1,280
RSOD (Long et al. 2017)	2017	TGRS	horizontal bounding box	976	4	6,950	~1,000
HRSC2016 (Liu et al. 2016)	2016	GRSL	oriented bounding box	1,061	26	2,976	~1,100
COWC (Mundhenk et al. 2016)	2016	ECCV	Center point	53	1	32,716	2,000–19,000
UCAS-AOD (Zhu et al. 2015)	2015	ICIP	oriented bounding box	1,510	2	14,596	~1,000
DLR 3k (Liu and Mattyus 2015)	2015	GRSL	oriented bounding box	20	8	14,235	5,616
VEDAI (Razakarivony and Jurie 2016)	2016	JVCI	oriented bounding box	1,268	9	2,950	**512,104**
NWPU VHR-10 (Cheng et al. 2014)	2014	ISPRS	horizontal bounding box	800	10	3,651	~1,000
SZTAKI-INRIA (Benedek et al. 2011)	2011	TPAMI	oriented bounding box	9	1	665	~800
TAS (Heitz and Koller 2008)	2008	ECCV	horizontal bounding box	30	1	1,319	792

2.3. DOTA dataset

2.3.1. DOTA-v1.0

To address the current demand for object detection research in RSIs, Xia et al. (2018) introduced the DOTA-v1.0 dataset. In previous datasets for ODAI, NWPU VHR-10 (Cheng et al. 2014) garnered attention due to its coverage of multiple categories. However, the DOTA-v1.0 dataset surpasses NWPU VHR-10 in terms of the number of categories and instances per category. Fig. 4 provides a comparative display of the category composition of the DOTA-v1.0 dataset and the NWPU VHR-10 dataset, with the horizontal axis representing target categories and the vertical axis indicating the number of categories.

Compared to previous datasets for ODAI, the DOTA-v1.0 dataset exhibits the following prominent features:

(1) Large-scale: The DOTA-v1.0 dataset is extensive, comprising 2,806 images and 188,282 instances. It encompasses 15 different object categories, with pixel dimensions ranging from 800 × 800 to 4,000 × 4,000, encompassing a wide variety of object instances in terms of size, orientation, and shape.

(2) Comprehensive annotation: Each image in the DOTA-v1.0 dataset has been meticulously annotated, providing researchers with abundant training and validation data, concurrently contributing to the enhancement of algorithm accuracy and performance.

(3) Abundant instances: Scale variations in geographical objects are often of significant importance, influenced by both the spatial resolution of sensors and the disparities in size among different object categories including variations within the same category. The DOTA-v1.0 dataset exhibits a wealth of instance variations, encompassing diverse perspectives, scales, and orientations, including scenarios with small objects and large buildings.

(4) A more precise annotation approach: Unlike datasets like DIOR (Li et al. 2019b) and xView (Lam et al. 2018), which employ HBB, the DOTA-v1.0 dataset utilizes Oriented Bounding Boxes (OBB) for annotation, allowing for a more accurate description of object shape and orientation characteristics. Fig. 5 illustrates annotated image examples from the DOTA-v1.0 dataset. Compared to traditional horizontal bounding boxes, OBB can better conform to the contours of objects, thereby providing finer localization information for object detection tasks.

While the DOTA-v1.0 dataset possesses unique advantages in the field of ODAI, it also comes with certain limitations. These limitations encompass (1) the absence of annotations for tiny objects (less than 10 pixels); (2) the dataset primarily originates from a single domain, i.e., Google Earth images; (3) the images are typically selected from specific

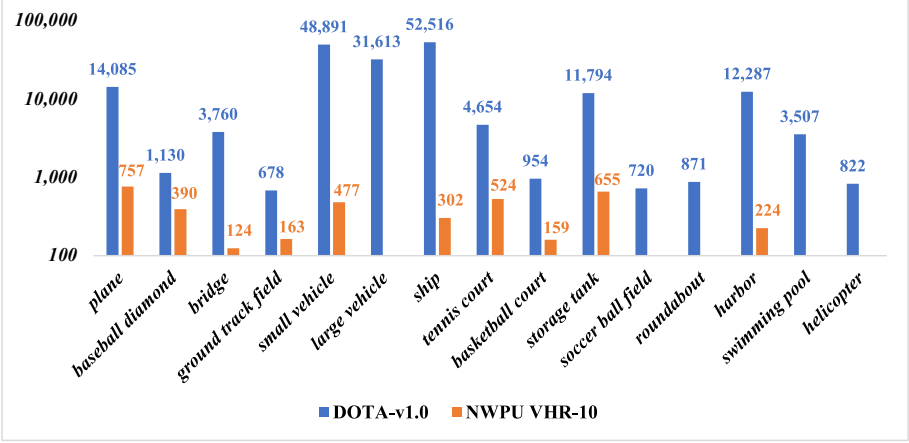


Fig. 4. Comparison between DOTA-v1.0 and NWPU VHR-10 in categories and responding quantity of instances.

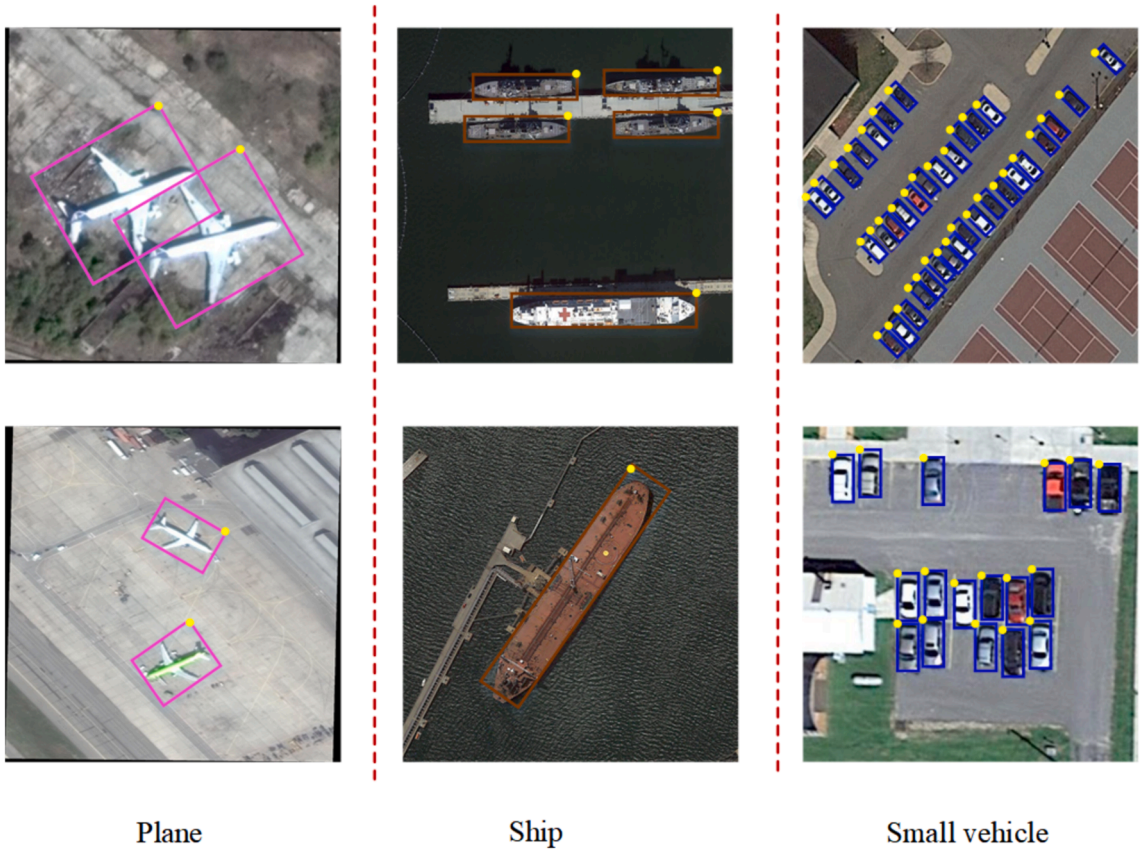


Fig. 5. Sample of annotated images in DOTA-v1.0.

regions within large-scale images that contain multiple objects.

These limitations may have certain implications for the generalization capabilities and practical applications of object detection algorithms. In recent years, the DOTA-v1.0 dataset has made significant strides in the field of ODAI, but there are still some challenging issues yet to be fully addressed. Firstly, benchmarking detection models for tiny and regular-sized oriented objects remains an unresolved challenge. Secondly, there are issues pertaining to large-scale image object detection, such as those containing only a few objects but with dimensions exceeding 20,000 × 20,000 pixels. Additionally, the development of object detection models suitable for multi-source high-altitude imagery to achieve robust detection across diverse targets still presents certain challenges in practical applications (Ding et al. 2021).

To address these limitations more effectively, the DOTA-v1.5 dataset has been expanded based on the DOTA-v1.0 dataset, specifically adding annotations for tiny objects and integrating data from various domains. Furthermore, the DOTA-v2.0 dataset has been further extended to include larger-sized GF-2 and airborne images, which align more closely with the distribution of targets in real-world applications. Interestingly, we found that all these versions use the same hyperparameters on the same detector.

2.3.2. DOTA-v1.5

Unlike the DOTA-v1.0 dataset, which primarily stems from a single domain, the DOTA-v1.5 dataset amalgamates image data from two different sources, i.e., Google Earth images and GAOFEN-2 and JILIN-1 (GF &JL) satellite images. In Table 4, we present statistics related to the image area, object area, and foreground ratio from these two sources, with some data cited from Ding et al. (2021). A deep dive into these statistics reveals that the meticulously curated Google Earth images encompass most of the positive samples. However, to counteract the potential bias towards positive samples, negative samples also play an indispensable role in balancing the sample distribution and enhancing model performance (Torralba and Efros 2011).

Table 4
The statistics for the annotated objects across different data sources in DOTA-v1.5.

	Google Earth	GF-2&JL-1	Total
Foreground Ratio	0.06	0.03	0.042
Image area (10^6 pixels)	11,873	7,301	19,173
Object area (10^6 pixels)	784	20	804
Image quantity	2,375	431	2,806

The DOTA-v1.5 dataset covers 402,089 instance objects. In terms of the number of images, it remains consistent with the DOTA-v1.0 dataset, with a total of 2,806 images. However, in terms of annotations, the DOTA-v1.5 dataset has been expanded to include instances that are extremely small (less than 10 pixels). Additionally, this dataset introduces a new category called "Container Crane". Similar to the split in the DOTA-v1.0 dataset, in DOTA-v1.5, 1,141 images are used for training, 458 images for validation, and 937 images for testing. Fig. 6 illustrates the number of target instances for each category in the DOTA-v1.5 dataset.

2.3.3. DOTA-v2.0

The DOTA-v2.0 dataset is derived from three primary sources, i.e., Google Earth images, GF-2 and JL-1 (GF&JF) satellite images, and the CycloMedia (Zhang et al. 2019c) airborne images. Table 5 presents the statistics for these three image sources, with some data referenced from the literature (Ding et al. 2021). By integrating GF-2 & JL-1 satellite imagery with CycloMedia RSIs, the DOTA-v2.0 dataset more accurately reflects the target distribution in real-world application scenarios and offers richer background information for models (Ding et al. 2021). It's

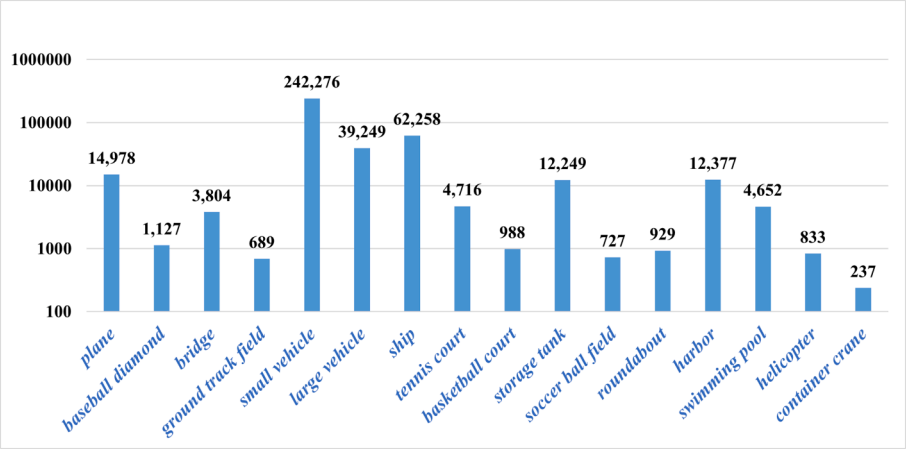


Fig. 6. The number of instances for each category in the DOTA-v1.5.

Table 5
The statistics for the annotated objects across different data sources in DOTA-v2.0.

	Google Earth	GF-2&JL-1	Aerial	Total
Foreground Ratio	0.037	0.003	0.033	0.016
Image area (10^6 pixels)	29,991	75,854	20,462	126,306
Object area (10^6 pixels)	1,111	243	673	2,027
Image quantity	10,186	516	566	11,268

worth noting that the DOTA-v2.0 dataset not only includes RGB images but also grayscale images. Specifically, images from Google Earth and CyloMedia are typically processed through RGB rendering to produce versions of the original RSIs, while GF-2 and JL-1 images achieve the creation of the original RSIs by converting each pixel's 8-bit original panchromatic band into a 10-bit representation. However, throughout these spectral rendering and bit-depth optimization processes, the structural and appearance-related information of the images remains consistent. Therefore, even after these processing steps, the images maintain good feasibility for recognition tasks (Ding et al. 2021).

The DOTA-v2.0 dataset encompasses 18 common categories, comprising a total of 11,268 images and 1,793,658 object instances. In comparison to the DOTA-v1.5 dataset, this version also introduces two new categories of "airport" and "helipad" (Ding et al. 2021). Based on our research, this is the largest publicly available Earth vision object detection dataset to date. The dataset is divided into training subsets, validation subsets, test-dev subsets, and test-challenge subsets. Specifically, the training subset contains 830 images and 268,627 instances, the validation subset includes 593 images and 81,048 instances, the test-dev subset comprises 2,792 images and 353,346 instances, and the test-challenge subset contains 6,053 images and 1,090,637 instances. Fig. 7 presents the number of object instances for each category in the DOTA-v2.0 dataset, where the horizontal axis represents object instance categories, and the vertical axis represents the number of object instances. Additionally, Fig. 8 showcases image examples from the training subset, validation subset, and test subset of the DOTA-v2.0 dataset.

In addition, we conducted a word cloud analysis of the titles of selected DOTA dataset-related papers, as shown in Fig. 9. From the word cloud, several high-frequency terms such as "Object Detection", "Remote Sensing", "Oriented", "Rotated", etc., are prominently visible. These high-frequency terms clearly reflect the widespread attention of the current ODAI community towards the DOTA dataset. These highlighted terms provide researchers with clear directions, helping them to pinpoint and focus on the current research areas more accurately, thus providing valuable guidance for achieving more targeted research outcomes.

3. Object detection in aerial images

In this section, we review traditional techniques for ODAI and methods rooted in deep learning.

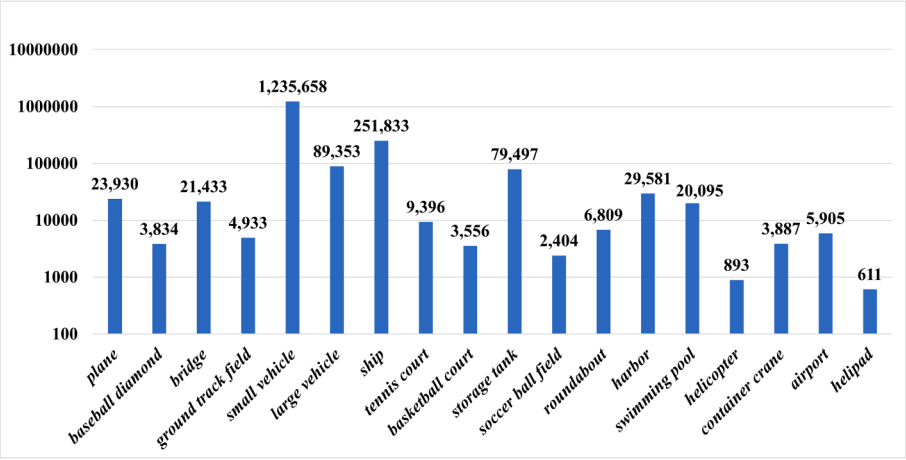


Fig. 7. The number of instances for each category in the DOTA-v2.0.

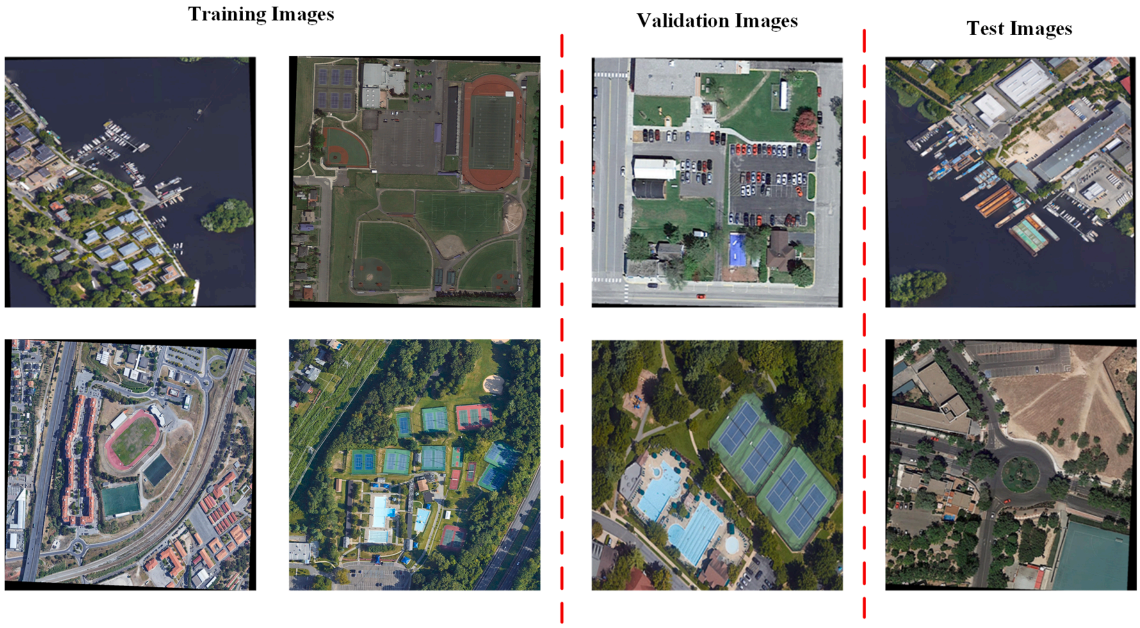


Fig. 8. The exhibition of DOTA-v2.0 dataset.

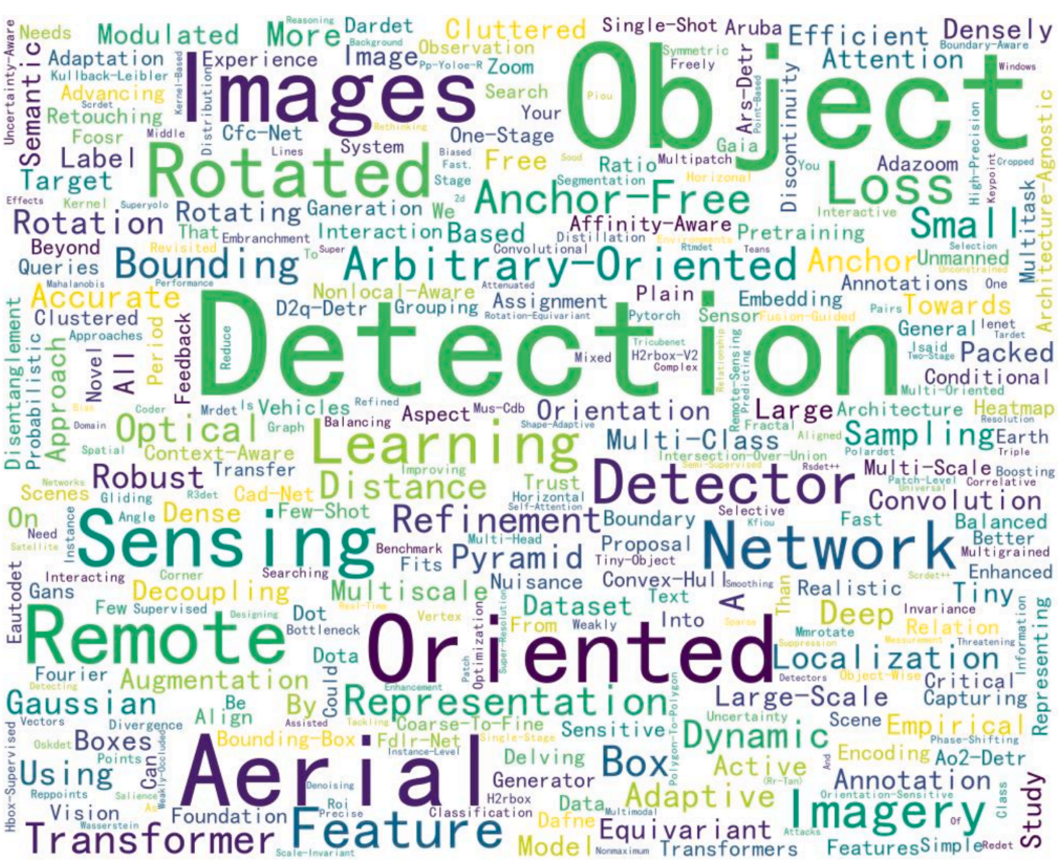


Fig. 9. Word cloud of paper titles reviewed by this paper.

3.1. Traditional methods for ODAI

Object detection methods in RSIs can be primarily divided into two categories: traditional methods and those based on deep learning. As shown in Fig. 10, traditional methods can be further classified into four subcategories: template matching-based methods, knowledge-based methods, object-based image analysis (OBIA) methods, and machine learning-based methods (Cheng and Han 2016). A qualitative analysis of these four subcategories, along with their respective strengths and weaknesses, is provided in Table 6.

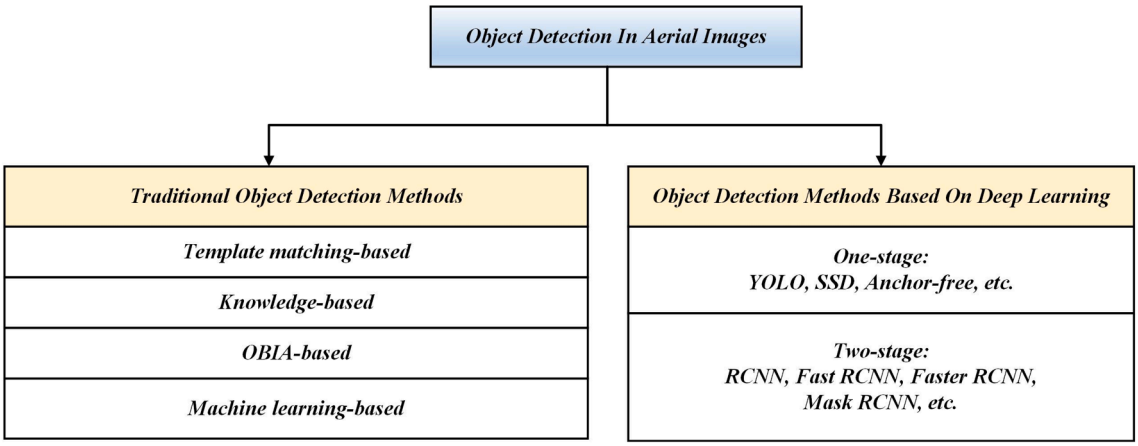


Fig. 10. Methods for object detection in RSIs.

3.2. Methods based on deep learning for object detection

In recent years, there have been significant breakthroughs in object detection methods based on deep learning. From the classic Region-based CNN (RCNN) (Girshick 2015) and You Only Look Once (YOLO) (Redmon et al. 2016) to more advanced approaches like RetinaNet (Lin et al. 2017a), these methods have not only brought new perspectives to the field of object detection but also provided crucial technical support for ODAI.

Deep Learning-based Object Detection Workflow typically can be divided into five main steps (Cheng and Han 2016), including data preprocessing, feature extraction and processing, bounding box generation, classification, and post-processing, as shown in Fig. 11. Inspired by the remarkable accomplishments of deep learning in the field of computer vision, the domain of ODAI has attracted the attention of numerous researchers in recent years. Over the past few decades, rich and diverse method have emerged for ODAI, which will be illustrated in details in Section 4.1.3.

4. Challenges and solutions for ODAI

In recent years, ODAI has attracted extensive attention, and numerous researchers have made significant contributions in this area. Table 7 presents the statistical results of the average detection accuracy (%) of cutting-edge methods using the DOTA-v1.0 dataset under similar model size and complexity budgets. The specific methods are: STD+HiViT-B (Yu et al. 2024), Large Selective Kernel Network (LSKNet-S*) (Li et al., 2023a), Real-Time Object Detectors (RTMDet-R-l) (Lyu et al. 2022), Rotated Varied-Size Attention (RVSA) (Wang et al. 2023b), KFIoU (Yang et al. 2023b), O-RCNN (Xie et al. 2021), Anchor-free Oriented Proposal Generator (AOPG) (Cheng et al. 2022a), Kullback-Leibler Divergence (R3Det-KLD) (Yang et al., 2021e), DODet (Cheng et al. 2022b), Gaussian Wasserstein distance (R3Det-GWD) (Yang et al. 2021d), Rotation-equivariant Detector (ReDet) (Han et al. 2021), Single-shot Alignment Network (S2A-Net) (Han et al. 2020), Arbitrary-Oriented Object Detection Transformer (AO2-DETR) (Dai et al. 2022b), spectra-aware screening mechanism (SASM) (Hou et al. 2022), dense one-stage anchor-free deep model (DAFNe) (Lang et al. 2021), Coupled-hypersphere-based Feature Adaptation (CFA) (Guo et al. 2021b), Refined Single-Stage Detector (R3Det) (Yang et al., 2019b), Circular Smooth Label (CSL) (Yang and Yan 2020), CenterMap (Wang et al. 2020), Gliding vertex (G.V.) (Xu et al. 2021c), RoI Transformer (RoI Trans.) (Ding et al. 2019), SCRDet (Yang et al. 2019c).

Fig. 12 offers a detailed depiction of the best mean Average Precision (mAP) models obtained annually since 2017 from research based on the DOTA-v1.0 dataset. DOTA-v1.0 dataset includes 15 categories, which are: plane, baseball diamond, bridge, ground track field, small vehicle, large vehicle, ship, tennis court, basketball court, storage tank, soccer ball field, roundabout, harbor, swimming pool, and helicopter. These 15 categories are represented by A1 to A15, respectively. The figure clearly demonstrates that, over time, there has been continuous progress in the field of ODAI. Understanding this developmental trend is crucial for grasping the evolution of this domain and guiding the progress of future detection methods. Detectors in this figure: Faster R-CNN trained on Oriented bounding boxes (FR-O) (Xia et al. 2018), Image Cascade Network (ICN) (Azimi et al. 2018), SCRDet (Yang et al. 2019c), Adaptive Period Embedding (APE) (Zhu et al. 2020), R3Det (Yang et al. 2019b), SCRDet++ (Yang et al. 2020b), S2A-Net (Han et al. 2020), GWD+R3Det (Yang et al. 2021d), KLD+R3Det (Yang et al., 2021e), Oriented RCNN (Xie et al. 2021), KFIoU+RoI Trans (Yang et al. 2023b), RTMDet-R-l (Lyu et al. 2022), LSKNet-S* (Li et al. 2023a), STD+HiViT-B (Yu et al. 2024).

However, ODAI still confronts a series of challenging issues. According to the literature we found, these challenges encompass but are not limited to small object detection, scale variation, arbitrary-oriented object detection, and dense object detection. Based on the characteristics and application needs of the target, ODAI can be broadly categorized into two primary types: oriented object detection and rotated object detection. Delving deeper, these two primary categories cover a variety of sub-tasks, including small object detection, multi-scale object detection, arbitrary-oriented object detection, and dense object detection, as depicted in Fig. 13. Given the unique nature of these sub-tasks, ODAI continues to pose formidable challenges.

4.1. Challenges and solutions

In this section, we review over 150 pieces of literature, delving deeply into the various challenges faced in ODAI and presenting solutions proposed for these challenges.

4.1.1. Oriented object detection

4.1.1.1. Small object detection. In addressing the issue of small objects often being overlooked or mislabeled, several studies have made significant contributions. Liang et al. (2022) introduced a dynamic enhancement anchor network (DEA-Net), designed to generate novel training samples and facilitate interactive sample selection between anchor-based and anchor-free units. Lee et al. (2022) introduced Class-wise Collated Cor-relation (C3Det), an interactive annotation method for multiple instances of tiny objects from multiple classes, proven to be 2.85 times faster than manual annotation. Additionally, Xu et al. (2023) proposed Dynamic Coarse-to-Fine Learning (DCFL), a dynamic prior along with the coarse-to-fine assigner, aimed at addressing mismatches and imbalances in small object detection.

Table 6
Traditional object detection methods.

	Method	Strengths	Limitations
Template matching-based	Rigid template (Lefèvre et al. 2007; Weber and Lefèvre 2012; Weber and Lefevre 2008; Kim et al. 2004)	Simple and easy to implement.	Rotation and scale dependent; Sensitive to viewpoint changes and shape.
	Deformable template (Lin et al. 2017b; Liu et al. 2012; Tao et al. 2010)	More flexible and powerful in handling shape deformations and intra-class variations than rigid shape matching.	Template design requires more geometric shape prior information and parameters; High computational cost.
Knowledge-based	Geometric knowledge (Chaudhuri and Samal 2008; Akcay and Aksoy 2010; Haala and Brenner 1999)	Detection can be performed through a hierarchical structure from coarse-to-fine.	How to define detection rules and prior knowledge is subjective; Overly loose rules can lead to false positives.
	Context knowledge (Janssen and Middelkoop 1992; Wang and Newkirk 1988; Moon et al. 2002)		
OBIA-based	Image segmentation (Chen et al. 2012; Dissanska et al. 2009; Blaschke 2003)	The flexible combination of texture, shape, geometric features, and contextual semantic features, along with functionalities similar to Geographic Information Systems and expert knowledge, enables OBIA to context-aware and support multiple sources.	A universal solution for fully automated segmentation is still lacking; The expert knowledge on how to define classification rules remains subjective.
	Object classification (Blanzieri and Melgani 2008; Feizizadeh et al. 2014; Stumpf and Kerle 2011)		
Machine learning-based	HOG, Haar-like, SVM, AdaBoost, KNN, etc. (Lei et al. 2011; Ari and Aksoy 2014; Benedek et al. 2015)	Detector can be automatically established using machine learning techniques; The detection system has scalability and compatibility; It possesses high detection accuracy.	The learning of the classifier requires a large number of object and non-object training samples; Detection precision is influenced by the training data.

In response to the lack of appearance information for tiny objects and the interference of complex backgrounds, innovative network frameworks (Doloriel and Cajote 2023; Zhang et al. 2022b; Zhang et al. 2023) were devised to extract richer features from tiny objects. Various researchers have designed various feature pyramid networks (Shamsolmoali et al. 2022b; Gu et al. 2022; Shamsolmoali et al. 2022a; Shamsolmoali et al. 2021; Azimi et al. 2018) to enhance the feature extraction capability for small-scale objects. Additionally, Xu et al. (2021a) introduced a novel Intersection over Union (IoU) metric called DotD for tiny object detection. Li et al. (2019a) introduced a method to learn novel object-wise semantic representation, which suppresses background interference and more accurately estimates proposals, thereby enhancing performance in small object detection within RSIs.

4.1.1.2. Multi-scale object detection. To address the scale variation issue in oriented object detection, different network models (Xu et al. 2021b; Yang et al. 2019a; Xiao et al. 2023) have been designed effectively handling the extreme scale variations present in RSIs. In addition, Zhao et al. (2021) introduced a method named PolarDet that employs polar coordinates to represent the oriented objects, enhancing the accuracy of multi-scale object classification. Meanwhile, Hou et al. (2022) presented a shape-adaptive selection (SA-S) and shape-adaptive measurement (SA-M) strategies tailored for oriented object detection, aiming to tackle the challenges of object shape variations in RSIs.

4.1.1.3. Arbitrary-oriented object detection. To tackle the challenges associated with arbitrary-oriented object detection, researchers have proposed a plethora of innovative approaches. The complexities of angular computation were overcome by optimizing label allocation strategies (Zhu et al. 2020; Ming et al. 2022a), thereby enhancing the accuracy of arbitrary-oriented object detection. Different object detection network frameworks using anchor-free strategies (Huang et al. 2022a; Lin et al. 2019; Chen et al. 2022; Cheng et al. 2022a; Lang et al. 2021; Li et al. 2023b; Wei et al. 2019; Wang et al. 2022b) were introduced by various researchers, leading to improved computational efficiency and performance in oriented object detection. By employing attention mechanism (Wang et al. 2023b; Yang et al. 2021a; Liu and Hu 2022; Zhang et al. 2019a) to extract richer contextual information, enhanced the object representation capabilities, addressing the challenges brought by objects in arbitrary orientations in RSIs. Various feature alignment networks (Han et al. 2020; Guo et al. 2021a) were devised by researchers to mitigate the inconsistency between localization precision and classification score. The accuracy of oriented object detectors (Huang et al. 2022b; Li et al. 2023a) was boosted by dynamically adjusting the receptive field range. Networks based on the Transformer architecture (Tang et al. 2022; Dai et al. 2022b; Zeng et al. 2023; Ding et al. 2019; Zhou et al. 2023; Ma et al. 2021) were proposed to elevate the effectiveness of arbitrary-oriented object detection. Finally, various novel approach of learning rotated boxes from more readily available horizontal boxes was presented (Yang et al., 2023a; Yu et al. 2023). By solely using horizontal box annotations for weakly-supervised training, they achieved results competitive with methods trained using rotated boxes.

Moreover, in addressing the challenges of large variations in arbitrary-oriented objects within RSIs, researchers have proposed other innovative methods. Xu et al. (2021c) introduced a multi-oriented object detection method that adjusts the vertices of horizontal bounding boxes by sliding along their corresponding edges. This precisely describes multi-oriented objects, addressing the inapplicability of horizontal bounding boxes in oriented object detection. Wang et al. (2022a) introduced a multi-grained angle representation (MGAR) method and an Intersection over Union (IoU)-aware FAR-loss (IFL). The MGAR method enhances angle prediction accuracy, and the IFL improves angle representation, effectively resolving issues related to angle representation and loss design. Yi et al. (2021) proposed a box boundary-aware vectors (BBAVectors) method that, by incorporating box boundary-aware vectors, enhances object detection performance and tackles the imbalance issue between positive and negative anchor boxes. Yu and Da (2023) introduced a novel PSC, designed for accurately predicting the orientation of objects. Kim et al. (2022) presented TricubeNet, a novel oriented object detection approach that addresses angle discontinuity and reduces computational complexity.

4.1.1.4. Dense object detection. The densely packed objects presents significant challenges for ODAI. Fang et al. (2022) introduced a novel network. This network dynamically determines feature enhancement weights by measuring the affinity between objects. It aims to leverage the sharing similar orientations of densely packed objects to enhance rotated features, thereby improving the detection performance. Guo et al. (2021b) proposed a novel CFA method that adapts convolutional features to the layout of densely packed objects. This approach seeks to

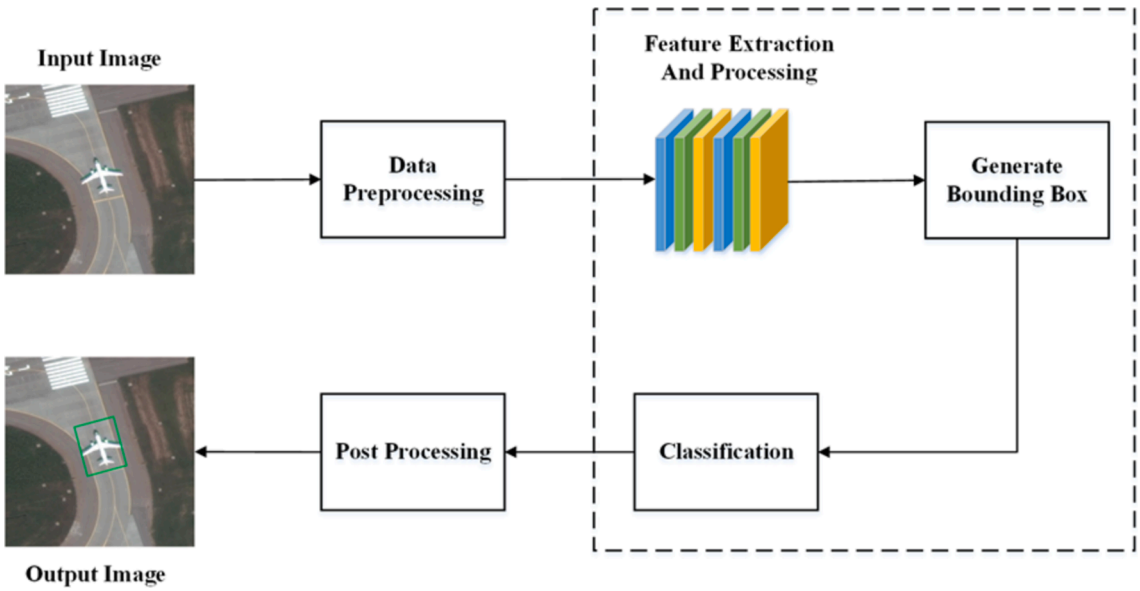


Fig. 11. The process of ODAI based on deep learning.

address the spatial feature aliasing caused by the intersection of reception fields between densely packed objects, ultimately increasing the accuracy of object detection. Pan et al. (2020) introduced a novel feature selection module (FSM) and dynamic refinement head (DRH). The FSM adjusts the receptive fields adaptively based on the shapes and orientations of targets, while the DRH enables the model to predict in an object-aware manner, addressing the challenges of detecting densely packed objects. Qin et al. (2020) proposed an arbitrary-oriented region proposal network (AO-RPN), generating transformed oriented proposals from horizontal anchors to learn features most suitable for specific tasks, achieving precise detection of densely packed arbitrary-oriented objects in RSIs.

Through the summary and analysis of these studies, we can observe that researchers have employed various innovative methods and network frameworks for small object, multi-scale, arbitrary-oriented, and dense object detection, aiming to enhance accuracy, efficiency, and robustness. These approaches have not only made theoretical breakthroughs but have also demonstrated significant performance improvements in practical applications, contributing positively to the development of target detection in remote sensing imagery. However, it is important to note that the applicability of different methods may vary across different scenarios and datasets. Further research is needed to validate the generalizability and practicality of these methods in the future.

4.1.2. Rotated object detection

4.1.2.1. Small object detection. In recent years, there have been significant advancements in the field of ODAI. However, challenges persist in detecting small size objects within RSIs. Addressing these tiny and densely packed objects, Yang et al. (2019c) introduced a novel multi-category rotation object detector called SCRDet, which notably improved the detection capabilities for such objects. Subsequently, Yang et al. (2020b) unveiled an enhanced version named SCRDet++. They first innovatively introduced the idea of denoising to object detection, implementing instance-level denoising on the feature maps, which further amplified the detection proficiency for small and densely packed objects.

4.1.2.2. Multi-scale object detection. Rotated object detection in RSIs is particularly challenging due to the significant variations of scale, aspect ratio, rotation, and densely packed targets. To address these challenges, numerous innovative methods have been introduced by researchers. For instance, various researchers proposed anchor-free rotated object detectors (Zhang et al. 2022a; Dai et al. 2022a) to handle scale variations in rotated object detection. Meanwhile, Zand et al. (2022) introduced an object detection method tailored for freely rotated objects of arbitrary sizes. Their approach harnesses classification learning to gather the necessary information, enabling precise detection of these objects.

4.1.2.3. Arbitrary-oriented object detection. The task of rotated object detection aims to identify and locate objects in arbitrary orientation within images. However, the vast variations in object orientations across different images, combined with the potential presence of multiple object classes within a single image, present considerable challenges for standard backbone networks to extract high-quality features from these arbitrary oriented objects. To address this, Pu et al. (2023) introduced an Adaptive Rotational Convolution (ARC) module. This module can effectively extract features from objects of varying orientations and adapt to a wide range of directional changes, alleviating the issue faced by standard backbone networks in extracting features from objects in arbitrary orientations. In addition, several rotated object detectors (Lu et al. 2022; Han et al. 2021) have been proposed to effectively address objects distributed in arbitrary orientation. Zhou et al. (2022) offered a coherent algorithm framework called MMRotate, further advancing research in the field of rotated object detection.

In RSIs of object detection, boundary discontinuity is a common and challenging issue. RSIs often feature complex backgrounds, low-contrast boundaries, and fine details, as well as noise interference. These factors make it difficult to accurately detect and continuously represent the boundaries of the targets. In the design of regression loss for rotated object detection, boundary discontinuity remains a central challenge. To fundamentally address this discontinuity issue in regression-based detectors, several innovative methods (Yang et al. 2021b; Ming et al. 2020; Yang et al. 2020a; Murrugarra-Llerena et al. 2021) have been proposed by researchers. These techniques aim to effectively tackle the regression problems in rotated object detection. Furthermore, in the pursuit of improving the precision of rotated object detection, an array of distinct loss functions (Yang et al., 2021e; Qian et al. 2019; Ming et al. 2022b; Yang et al. 2022c; Yang et al. 2021d; Qian et al. 2022; Wen et al. 2022; Chen et al. 2020b; Sun et al. 2018; Yang et al. 2023b) have been introduced by various researchers. These loss functions are designed to effectively address the boundary regression challenges in rotational

Table 7
Statistics of the average detection accuracy (%) of state-of-the-art methods on the DOTA-v1.0 dataset.

A1 plane A9 basketball court		A2 baseball diamond A10 storage tank		A3 bridge A11 soccer ball field		A4 ground track field A12 roundabout		A5 small vehicle A13 harbor		A6 large vehicle A14 swimming pool		A7 ship A15 helicopter		A8 tennis court — —	
Method	**year**	**mAP↓**	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**	**A12**	**A13**	**A14**	**A15**
STD+HiViT-B	2024	**82.24**	89.15	85.03	60.79	82.06	80.90	85.76	88.45	90.83	87.71	87.29	**73.99**	71.25	**85.18**	82.17	**82.95**
LSKNet-S*	2023	81.85	89.69	85.70	61.47	**83.23**	81.37	86.05	88.64	90.88	88.49	87.40	71.67	71.35	79.19	81.77	80.86
RTMDet-R-l	2022	81.33	88.01	86.17	58.54	82.44	81.30	84.82	88.71	90.89	**88.77**	87.37	71.96	71.18	81.23	81.40	77.13
RVSA	2022	81.24	88.97	85.76	61.46	81.27	79.98	85.31	88.30	90.84	85.06	87.50	66.77	**73.11**	84.75	81.88	77.58
KFIoU	2022	80.93	89.44	84.41	**62.22**	82.51	80.10	**86.07**	88.68	**90.90**	87.32	88.38	72.80	71.95	78.96	74.95	75.27
O-RCNN	2021	80.87	89.84	85.43	61.09	79.82	79.71	85.35	88.82	90.88	86.68	87.73	72.21	70.80	82.42	78.18	74.11
AOPG	2022	80.66	89.88	85.57	60.90	81.51	78.70	85.29	88.85	90.89	87.60	87.65	71.66	68.69	82.31	77.32	73.10
R3Det-KLD	2021	80.63	89.92	85.13	59.19	81.33	78.82	84.38	87.50	89.80	87.33	87.00	72.57	71.35	77.12	79.34	78.68
DODet	2022	80.62	89.96	85.52	58.01	81.22	78.71	85.46	88.59	90.89	87.12	87.80	70.50	71.54	82.06	77.43	74.47
R3Det-GWD	2021	80.23	89.66	84.99	59.26	82.19	78.97	84.83	87.70	90.21	86.54	86.85	73.47	67.77	76.92	79.22	74.92
ReDet	2021	80.10	88.81	82.48	60.83	80.82	78.34	86.06	88.31	90.87	**88.77**	87.03	68.65	66.90	79.26	79.71	74.67
S2A-Net	2020	79.42	88.89	83.60	57.74	81.95	79.94	83.19	**89.11**	90.78	84.87	87.81	70.30	68.25	78.30	77.01	69.58
AO2-DETR	2022	79.22	89.95	84.52	56.90	74.83	80.86	83.47	88.47	90.87	86.12	**88.55**	63.21	65.09	79.09	**82.88**	73.46
SASM	2022	79.17	89.54	85.94	57.73	78.41	79.78	84.19	89.25	90.87	58.80	87.27	63.82	67.81	78.67	79.35	69.37
DAFNe	2021	76.95	89.40	**86.27**	53.7	60.51	**82.04**	81.17	88.66	90.37	83.81	87.27	53.93	69.38	75.61	81.26	70.86
CFA	2021	76.67	89.08	83.20	54.37	66.87	81.23	80.96	87.17	90.21	84.32	86.09	52.34	69.94	75.52	80.76	67.96
R3Det	2021	76.47	89.8	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.57	62.68	67.53	78.56	72.62
CSL	2020	76.17	**90.25**	85.53	54.64	75.31	70.44	73.51	77.62	90.84	86.15	86.69	69.60	68.04	73.83	71.10	68.93
CenterMap	2020	76.03	89.83	84.41	54.60	70.25	77.66	78.32	87.19	90.66	84.89	85.27	56.46	69.23	74.13	71.56	66.06
G.V.	2021	75.02	89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32
RoI Trans.	2019	74.61	88.65	82.6	52.53	70.87	77.93	76.67	86.87	90.71	83.83	82.51	53.95	67.61	74.67	68.75	61.03
SCRDet	2019	72.61	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21

object detection.

4.1.2.4. Dense object detection. In RSIs, the challenges of accurately locating multi-angle objects and effectively distinguishing them from the background make rotated object detection a formidable task. Despite considerable progress, practical settings still present issues with densely distributed and rotated objects, and extreme class imbalance. These factors further complicate the detection task. To address these challenges, Yang et al. (2021c) introduced a novel single-stage rotation detector, which employs a progressive regression approach from coarse to fine granularity, ensuring fast and accurate object detection.

Overall, these studies indicate continuous innovation in the ODAI field to address diverse challenges that may arise in different scenarios. The introduction of various methods and technologies has laid a solid foundation for the development of this field, providing insights for future research and further innovation.

4.1.3. Other challenges

In terms of addressing data augmentation and sample imbalance issues, researchers have proposed various methods. Milz et al. (2018) introduced a realistic data augmentation technique based on conditional generative adversarial network (cGAN) for generating multi-sensor datasets. Sairam et al. (2023) proposed an architecture-agnostic balanced loss function called ARUBA to tackle the issue of size imbalance in drone-based aerial image datasets. Waqas Zamir et al. (2019) created a large-scale and densely annotated instance segmentation dataset (iSAID) tailored for remote sensing image instance segmentation tasks. Hong et al. (2019) introduced a method called hard chip mining, ensuring a balanced ratio across categories and generating hard examples that are effective for model training.

In terms of pretraining and transfer learning issues, Wang et al. (2023a) conducted pretraining on various networks using the current largest remote sensing scene recognition dataset, MillionAID, to enhance the fine-tuning performance of deep models in aerial scene tasks. Bu et al. (2021) introduced the GAIA transfer learning system, which can automatically and efficiently generate solutions tailored for heterogeneous downstream task requirements.

In terms of model design and improvement, researchers designed various object detection network architectures to enhance the performance of ODAI. To address significant limitations such as appearance occlusion and target size variation in aerial images, Shen et al. (2023) explored the limitations of traditional neck networks in object detection by analyzing the information bottleneck. They proposed an improved neck network to solve the problem of insufficient information in the current neck networks. To most effectively train a model on a large heterogeneous dataset, previous methods have typically employed separate detection heads on a shared backbone. However, this approach often results in a significant increase in parameters. Jain et al. (2024) proposed Mixture-of-Experts (MoE) as a solution, emphasizing that MoE is not just a scalability tool. By learning to route each dataset token to its corresponding expert, MoE trains these experts to become specialized for specific datasets.

Moreover, to address the significant differences in the number and size of objects between aerial images and consumer data, Biswas and Tešić (2024) proposed a compact object detection pipeline. This pipeline improves the feature extraction process using a spatial pyramid pooling, a cross-stage partial network, and a heatmap-based region proposal network (RPN). Due to the significant variation in object scales and different ranging environments in remote sensing images, previous methods have used large kernel convolutions or dilated convolutions. However, the former often introduces considerable background noise, while the latter may lead to overly sparse feature representations. Cai et al. (2024) proposed PKINet, a network that employs multi-scale convolutional kernels, allowing it to extract features of objects at different scales and capture local context without the need for dilation.

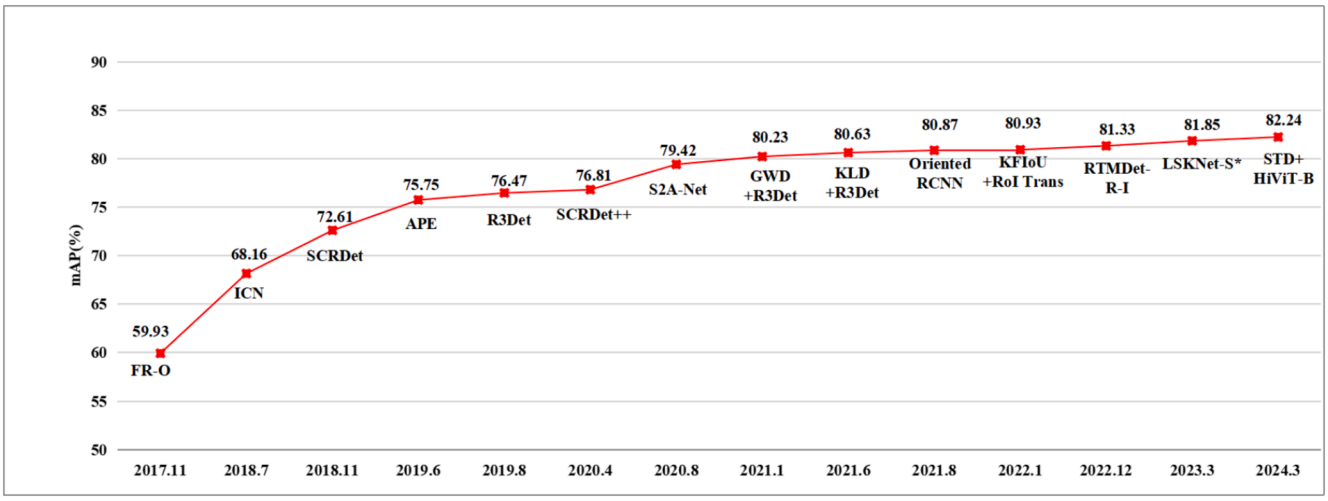


Fig. 12. The models with the highest mAP in ODAI on the DOTA-v1.0 dataset each year since 2017.

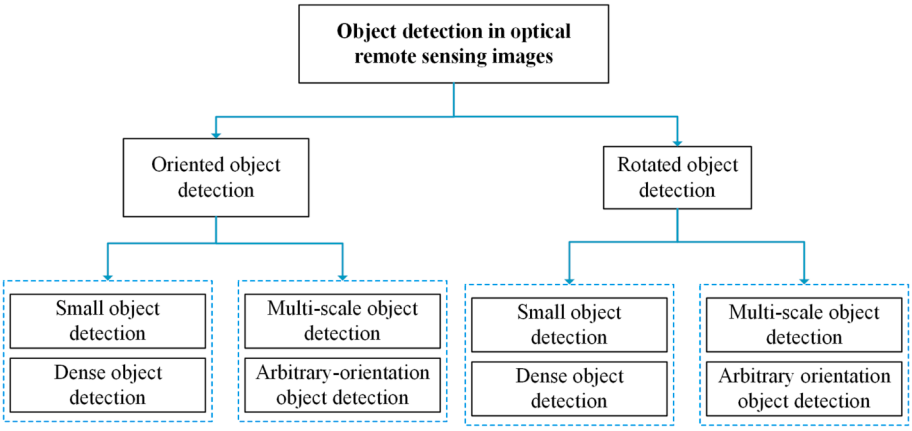


Fig. 13. Object detection in optical RSIs.

Current semi-supervised object detection (SSOD) methods primarily focus on detecting horizontal objects, with relatively little research dedicated to detecting objects in arbitrary orientations within remote sensing images. Inspired by this limitation, Fu et al. (2024) proposed a semi-supervised object detection framework (S^2O-Det) aimed at reducing annotation costs while improving detection performance in a semi-supervised manner.

Moreover, since semi-supervised oriented object detection (SSOD) has not been explored in RSIs, Hua et al. (2023) innovatively proposed a semi-supervised oriented object detection model built upon the mainstream pseudo-labeling framework for ODAI.

In the field of ODAI, researchers have proposed several other innovative methods. Jeune et al. (2021) introduced a few-shot representation learning method for ODAI, achieving online adaptation to new categories, providing stronger adaptability for handling different classes of targets. Researchers introduced various novel active learning methods (Qu et al. 2020; Liang et al. 2023), aiding in addressing the challenge of extensive manual annotation in ODAI. Furthermore, due to the presence of dense small targets in RSIs, traditional active learning approaches are inadequate. Hence, Li et al. (2022) proposed an adaptive points learning method, leveraging the benefits of adaptive points representation to capture the geometric information of objects in arbitrary orientation, addressing the detection of non-axis aligned targets in RSIs. Zheng et al. (2023) introduced a new localization distillation (LD) method, effectively transferring the teacher model's localization knowledge to the student model. This method introduced the concept of valuable localization region, selectively transmitting classification and localization knowledge. Varga and Zell (2021) presented a straightforward and effective slicing technique to enhance drone image detection performance. Chen et al. (2021) introduced an object detection method called Target Enhancement and Attenuated Nonmaximum Suppression (TEANS). This method employs a target enhancement framework to precisely detect objects, especially small ones, while using a weakened Nonmaximum Suppression (NMS) technique to overcome issues of erroneously suppressing dense target proposals. Shermeyer and Van Etten (2019) enhanced satellite images beyond their inherent resolution and investigated the impact of super-resolution techniques on object detection algorithm performance. Although there are increasingly general representation methods in the visual recognition field, research in object detection remains sparse. Wang et al. (2019) developed an effective and universal object detection system suitable for various image domains, including faces, traffic signs, and medical Computed Tomography (CT) images.

In summary, these studies collectively contribute to the continuous advancement of ODAI by addressing diverse challenges through innovative approaches in data handling, model training, and active learning. The variety of techniques presented provides a comprehensive understanding of the evolving landscape in ODAI, paving the way for future research and advancements in the field.

5. Future development directions

While deep learning techniques have made significant advancements in ODAI and have become increasingly mature, there are still certain shortcomings that merit further discussion. In this section, we delve into the existing challenges within ODAI and shed light on potential future research trends.

Large-scale models that support multi-resolution and cross-platform applications: Different RSIs often have varied pixel resolutions. This diversity in resolution poses challenges for object detection across these images. While large-scale deep learning models excel in handling high-resolution images, they also demand extensive and high-quality datasets. Consequently, constructing datasets with a range of resolutions to cater to the training needs of these large models could be a promising avenue for exploration.

Weakly supervised learning for rotation detectors: Weakly supervised object detection aims to reduce reliance on data annotation by using image-level annotations instead of bounding box annotation (Zhang et al. 2021). From our research, weakly supervised rotation detectors have not been created because there is no bounding box annotation information available. Also, the horizontal bounding boxes used in current object detection models fail to accurately locate objects in complexly oriented remote sensing imagery. Therefore, exploring the application of weakly supervised learning in rotated object detection could be a promising direction.

Developing techniques to adapt the foundation models to ODAI: With the recent advances in foundation models such as SAM (Segment Anything) and CLIP, there is an increasing focus on developing techniques to adapt these models to ODAI (Lacoste et al. 2024; Wang et al. 2024). We believe that developing methods to adapt foundation models to ODAI is a key trend for future research. This includes enhancing the models' ability to accurately detect and classify objects in the complex and varied contexts typical of aerial imagery. Future directions will likely involve the integration of multimodal data sources, such as combining optical images with LiDAR and infrared data, to improve the robustness and precision of detections. Additionally, there will be a significant emphasis on fine-tuning and transfer learning to customize these foundation models for the specific characteristics and challenges of aerial imagery. The creation of extensive, annotated datasets specific to ODAI will be essential in this process. Furthermore, advancing real-time processing capabilities to meet the demands of applications like surveillance, disaster response, and urban planning will be critical. Ultimately, the successful adaptation of foundation models to ODAI has the potential to significantly advance the field, leading to more accurate, reliable, and versatile detection systems.

6. Conclusions

This article provides a detailed review of research advancements made on the DOTA dataset over the past six years, covering over 150 papers. Initially, we introduce several standard datasets for natural scene image object detection and ODAI. Utilizing charts, we perform an in-depth comparative analysis between them and the DOTA dataset in terms of image count, annotation methods, pixel area, etc., illustrating the differences between natural scene images and RSIs, while also highlighting the unique advantages of the DOTA dataset. Subsequently, we showcase three versions of the DOTA dataset, detailing the strengths and limitations of the initial DOTA-v1.0 version and the enhancements brought by the subsequent DOTA-v1.5 and DOTA-v2.0 versions. We also revisit traditional object detection methods in RSIs, summarizing the pros and cons of various methods, and delve into deep learning-based object detection techniques. To give readers a clear understanding of the latest research developments achieved with the DOTA dataset in the domain of ODAI, we have selected and summarized approximately 110 research papers highly relevant to the DOTA dataset.

In conclusion, this paper has explored the current challenges in the domain of ODAI. Key findings indicate that the establishment of datasets tailored for extensive model training, the integration of weakly supervised learning in rotating object detection, and the advancement of lightweight network models stand as pivotal research directions and opportunities for future endeavors in this domain.

CRediT authorship contribution statement

Ziyi Chen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Huayou Wang:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xinyuan Wu:** Validation, Software, Methodology, Conceptualization. **Jing Wang:** Writing – review & editing, Resources. **Xinrui Lin:** Writing – review & editing, Supervision, Formal analysis. **Cheng Wang:** Supervision, Conceptualization. **Kyle Gao:** Writing – review & editing. **Michael Chapman:** Writing – review & editing. **Dilong Li:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by Natural Science Foundation of Fujian Province (No. 2023J01135), National Natural Science Foundation of China (No. 62001175), Fundamental Research Funds for the Central Universities of Huaqiao University (No. ZQN-911), the National Natural Science Foundation of China (No. 42201475), and the Natural Science Foundation of Fujian Province (No. 2021J05059).

References

Akcay, H.G., Aksoy, S., 2010. Building detection using directional spatial constraints. In: Proc. IEEE Int. Geosci. Remote Sens. Symp., pp. 1932–1935. https://doi.org/10.1109/IGARSS.2010.5652842.

Ari, C., Aksoy, S., 2014. Detection of compound structures using a Gaussian mixture model with spectral and spatial constraints. IEEE Trans. Geosci. Remote Sens. 52 (10), 6627–6638. https://doi.org/10.1109/TGRS.2014.2299540.

Azimi, S.M., Vig, E., Bahmanyar, R., Körner, M., Reinartz, P., 2018. Towards multi-class object detection in unconstrained remote sensing imagery. In: Proc. Asian Conf. Comput. vis., pp. 150–165. https://doi.org/10.1007/978-3-030-20893-6_10.

Benedek, C., Descombes, X., Zerubia, J., 2011. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34, 33–50. https://doi.org/10.1109/TPAMI.2011.94.

Benedek, C., Shadaydeh, M., Kato, Z., Szirányi, T., Zerubia, J., 2015. Multilayer Markov random field models for change detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 107, 22–37. https://doi.org/10.1016/j.isprsjprs.2015.02.006.

Biswas, D., Tešić, J., 2024. Domain adaptation with contrastive learning for object detection in satellite imagery. IEEE Trans. Geosci. Remote Sens. 62, 1932–1935. https://doi.org/10.1109/TGRS.2024.3391621.

Blanzieri, E., Melgani, F., 2008. Nearest neighbor classification of remote sensing images with the maximal margin principle. IEEE Trans. Geosci. Remote Sens. 46 (6), 1804–1811. https://doi.org/10.1109/TGRS.2008.916090.

Blaschke, T., 2003. Object-based contextual image classification built on image segmentation. In: Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data. pp. 113-119. https://doi.org/10.1109/WARSD.2003.1295182.

Bu, X., Peng, J., Yan, J., Tan, T., Zhang, Z., 2021. GAIA: A Transfer Learning System of Object Detection that Fits Your Needs. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 274–283. https://doi.org/10.1109/CVPR46437.2021.00034.

Cai, X., Lai, Q., Wang, Y., Wang, W., Sun, Z., Yao, Y., 2024. Poly kernel inception network for remote sensing detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 27706–27716. https://doi.org/10.48550/arXiv.2403.06258.
Chaudhuri, D., Samal, A., 2008. An automatic bridge detection technique for multispectral images. IEEE Trans. Geosci. Remote Sens. 46 (9), 2720–2727. https://doi.org/10.1109/TGRS.2008.923631.
Chen, Z., Chen, K., Lin, W., See, J., Yu, H., Ke, Y., Yang, C., 2020b. Oriented object detection by searching corner points in remote sensing imagery. IEEE Geosci. Remote. Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3079314.
Chen, K., Wu, M., Liu, J., Zhang, C., 2020a. FGSD: A dataset for fine-grained ship detection in high resolution satellite images. https://doi.org/10.48550/arXiv.2003.06832.
Chen, G., Hay, G.J., Carvalho, L.M., Wulder, M.A., 2012. Object-based change detection. Int. J. Remote Sens. 33 (14), 4434–4457. https://doi.org/10.1080/01431161.2011.648285.
Chen, H.-B., Jiang, S., He, G., Zhang, B., Yu, H., 2021. TEANS: a target enhancement and attenuated nonmaximum suppression object detector for remote sensing images. IEEE Geosci. Remote. Sens. Lett. 18, 632–636. https://doi.org/10.1109/LGRS.2020.2983070.
Chen, X., Ma, L., Du, Q., 2022. Oriented object detection by searching corner points in remote sensing imagery. IEEE Geosci. Remote. Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3079314.
Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 117, 11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014.
Cheng, G., Han, J., Zhou, P., Guo, L., 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS J. Photogramm. Remote Sens. 98, 119–132. https://doi.org/10.1016/j.isprsjprs.2014.10.002.
Cheng, G., Wang, J., Li, K., Xie, X., Lang, C., Yao, Y., Han, J., 2022a. Anchor-free oriented proposal generator for object detection. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2022.3183022.
Cheng, G., Yao, Y., Li, S., Li, K., Xie, X., Wang, J., Yao, X., Han, J., 2022b. Dual-aligned oriented detector. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2022.3149780.
Dai, L., Chen, H., Li, Y., Kong, C., Fan, Z., Lu, J., Chen, X., 2022a. TARDet: two-stage anchor-free rotating object detector in aerial images. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog. Workshops., pp. 4267–4275. https://doi.org/10.1109/CVPRW56347.2022.00472.
Dai, L., Liu, H., Tang, H., Wu, Z., Song, P., 2022b. Ao2-detr: Arbitrary-oriented object detection transformer. IEEE Trans. Circuits Syst. Video Technol. 33, 2342–2356. https://doi.org/10.1109/TCSVT.2022.3222906.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
Ding, J., Xue, N., Long, Y., Xia, G.-S., Lu, Q., 2019. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 2844–2853. https://doi.org/10.1109/CVPR.2019.00296.
Ding, J., Xue, N., Xia, G., Bai, X., Yang, W., Yang, M.Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2021. Object detection in aerial images: a large-scale benchmark and challenges. IEEE Trans. Pattern Anal. Mach. Intell. 44, 7778–7796. https://doi.org/10.1109/TPAMI.2021.3117983.
Dissanska, M., Bernier, M., Payette, S., 2009. Object-based classification of very high resolution panchromatic images for evaluating recent change in the structure of patterned peatlands. Can. J. Remote Sens. 35 (2), 189–215. https://doi.org/10.5589/m09-002.
Doloriel, C.T.C., Cajote, R.D., 2023. Improving the Detection of Small Oriented Objects in Aerial Images. In: Proc. IEEE Winter Conf. Appl. Comput. Vis. Workshops. pp. 176-185. https://doi.org/10.1109/WACVW58289.2023.00023.
Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., Tian, Q., 2018. The unmanned aerial vehicle benchmark: Object detection and tracking. https://doi.org/10.48550/arXiv.1804.00518.
Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. Int. J. Comput. vis. 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
Everingham, M., Eslami, S.M.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. Int. J. Comput. vis. 111, 98–136. https://doi.org/10.1007/s11263-014-0733-5.
Fang, T., Liu, B., Zhao, Z., Chu, Q., Yu, N., 2022. Affinity-Aware Relation Network for Oriented Object Detection in Aerial Images. Proc. Asian Conf. Comput. vis. 13845, 3343–3360. https://doi.org/10.1007/978-3-031-26348-4_22.
Feizizadeh, B., Tiede, D., Moghaddam, M.R., Blaschke, T., 2014. Systematic evaluation of fuzzy operators for object-based landslide mapping. South-Eastern Eur. J. Earth Observ. Geomatics. 3 (2s), 219–222.
Fu, R., Yan, S., Chen, C., Wang, X., Heidari, A.A., Li, J., Chen, H., 2024. S^2O-Det: a semi supervised oriented object detection network for remote sensing images. IEEE Trans. Industr. Inform. https://doi.org/10.1109/TII.2024.3403260.
Girshick, R., 2015. Fast r-cnn. In: Proc. IEEE Int. Conf. Comput. vis., pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
Gu, L., Popov, E., Ge, D., 2022. Fast Fourier Convolution Based Remote Sensor Image Object Detection for Earth Observation. https://doi.org/10.48550/arXiv.2209.00551.
Guo, G., Fang, L., Yue, J., 2021a. Oriented Spatial Correlative Aligned Feature for Remote Sensing Object Detection. In: Proc. IEEE Int. Geosci. Remote Sens. Symp., pp. 5319–5322. https://doi.org/10.1109/IGARSS47720.2021.9554246.
Guo, Z., Liu, C., Zhang, X., Jiao, J., Ji, X., Ye, Q., 2021b. Beyond Bounding-Box: Convex-hull Feature Adaptation for Oriented and Densely Packed Object Detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 8788–8797. https://doi.org/10.1109/CVPR46437.2021.00868.
Gupta, A., Dollar, P., Girshick, R., 2019. Lvis: A dataset for large vocabulary instance segmentation. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 5356–5364. https://doi.org/10.1109/CVPR.2019.00550.
Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS J. Photogramm. Remote Sens. 54 (2–3), 130–137. https://doi.org/10.1016/S0924-2716(99)00010-6.
Han, J., Ding, J., Li, J., Xia, G.-S., 2020. Align deep features for oriented object detection. IEEE Trans. Geosci. Remote Sens. 60, 1–11. https://doi.org/10.1109/TGRS.2021.3062048.
Han, J., Ding, J., Xue, N., Xia, G.-S., 2021. ReDet: a rotation-equivariant detector for aerial object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 2785–2794. https://doi.org/10.1109/CVPR46437.2021.00281.
Heitz, G., Koller, D., 2008. Learning spatial context: Using stuff to find things. In: Proc. Eur. Conf. Comput. vis., pp. 30–43. https://doi.org/10.1007/978-3-540-88682-2_4.
Hong, S., Kang, S., Cho, D., 2019. Patch-level augmentation for object detection in aerial images. In: Proc. IEEE Int. Conf. Comput. vis. Workshops., pp. 127–134. https://doi.org/10.1109/ICCVW.2019.00021.
Hou, L., Lu, K., Xue, J., Li, Y., 2022. Shape-adaptive selection and measurement for oriented object detection. In: Proc. AAAI Conf. Artif. Intell., pp. 923–932. https://doi.org/10.1609/aaai.v36i1.19975.
Hsieh, M.-R., Lin, Y.-L., Hsu, W.H., 2017. Drone-based object counting by spatially regularized regional proposal network. In: Proc. IEEE Int. Conf. Comput. vis., pp. 4145–4153. https://doi.org/10.1109/ICCV.2017.446.
Hua, W., Liang, D., Li, J., Liu, X., Zou, Z., Ye, X., Bai, X., 2023. SOOD: towards semi-supervised oriented object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 15558–15567. https://doi.org/10.1109/CVPR52729.2023.01493.
Huang, Z., Li, W., Xia, X.-G., Tao, R., 2022a. A general Gaussian heatmap label assignment for arbitrary-oriented object detection. IEEE Trans. Image Process. 31, 1895–1910. https://doi.org/10.1109/TIP.2022.3148874.
Huang, Z., Li, W., Xia, X.-G., Wu, X., Cai, Z., Tao, R., 2022b. A Novel nonlocal-aware pyramid and multiscale multitask refinement detector for object detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–20. https://doi.org/10.1109/TGRS.2021.3059450.
Jain, Y., Behl, H., Kira, Z., Vineet, V., 2024. DAMEX: Dataset-aware Mixture-of-Experts for visual understanding of mixture-of-datasets. In: Proc. Adv. Neural Inf. Process. Syst.. https://doi.org/10.48550/arXiv.2311.04894.
Janssen, L.L., Middelkoop, H., 1992. Knowledge-based crop classification of a Landsat Thematic Mapper image. Int. J. Remote Sens. 13 (15), 2827–2837. https://doi.org/10.1080/01431169208904084.
Jeune, P.L., Lebbah, M., Mokraoui, A., Azzag, H., 2021. Experience feedback using Representation Learning for Few-Shot Object Detection on Aerial Images. In: Proc. Int. Conf. Mach. Learn. Appl. pp. 662-667. https://doi.org/10.1109/ICMLA52953.2021.00110.
Kim, B., Lee, J., Lee, S., Kim, D., Kim, J., 2022. TricubeNet: 2D kernel-based object representation for weakly-occluded oriented object detection. In: Proc. IEEE Winter Conf. Appl. Comput. Vis. pp. 167-176. https://doi.org/10.1109/WACV51458.2022.00348.
Kim, T., Park, S.R., Kim, M.G., Jeong, S., Kim, K.O., 2004. Tracking road centerlines from high resolution remote sensing images by least squares correlation matching. Photogramm. Eng. Rem. S. 70 (12), 1417–1422. https://doi.org/10.14358/PERS.70.12.1417.
Lacoste, A., Lehmann, N., Rodriguez, P., Sherwin, E., Kerner, H., Lütjens, B., Irvin, J., Dao, D., Alemohammad, H., Drouin, A., Gunturkun, M., 2024. Geo-bench: Toward foundation models for earth monitoring. In: Proc. Adv. Neural Inf. Process. Syst., p. 36. https://doi.org/10.48550/arXiv.2306.03831.
Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., McCord, B., 2018. xview: Objects in context in overhead imagery. https://doi.org/10.48550/arXiv.1802.07856.
Lang, S., Ventola, F.G., Kersting, K., 2021. DAFNe A one-stage anchor-free approach for oriented object detection. https://doi.org/10.48550/arXiv.2109.06148.
Lee, C., Park, S., Song, H., Ryu, J., Kim, S., Kim, H., Pereira, S., Yoo, D., 2022. Interactive Multi-Class Tiny-Object Detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 14116–14125. https://doi.org/10.1109/CVPR52688.2022.01374.
Lefèvre, S., Weber, J., Sheeren, D., 2007. Automatic building extraction in VHR images using advanced morphological operators. In: Proc. Urban Remote Sens. Joint Event., pp. 1–5. https://doi.org/10.1109/URS.2007.371825.
Lei, Z., Fang, T., Huo, H., Li, D., 2011. Rotation-invariant object detection of remotely sensed images based on texton forest and hough voting. IEEE Trans. Geosci. Remote Sens. 50 (4), 1206–1217. https://doi.org/10.1109/TGRS.2011.2166966.
Li, W., Chen, Y., Hu, K., Zhu, J., 2022. Oriented reppoints for aerial object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 1819–1828. https://doi.org/10.1109/CVPR52688.2022.00187.
Li, Y., Hou, Q., Zheng, Z., Cheng, M., Yang, J., Li, X., 2023a. Large selective kernel network for remote sensing object detection. In: Proc. IEEE Int. Conf. Comput. vis., pp. 16748–16759. https://doi.org/10.1109/ICCV51070.2023.01540.
Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2019b. Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS J. Photogramm. Remote Sens. 159, 296–307. https://doi.org/10.1016/j.isprsjprs.2019.11.023.
Li, C., Xu, C., Cui, Z., Wang, D., Jie, Z., Zhang, T., Yang, J., 2019a. Learning Object-Wise Semantic Representation for Detection in Remote Sensing Imagery. Proc. IEEE Conf. Comput. vis. Pattern. Recog. Workshops..

Liang, D., Geng, Q., Wei, Z., Vorontsov, D.A., Kim, E.L., Wei, M., Zhou, H., 2022. Anchor Retouching via Model Interaction for Robust Object Detection in Aerial Images. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2021.3136350.

Liang, D., Zhang, J.-W., Tang, Y.-P., Huang, S.-J., 2023. MUS-CDB: mixed uncertainty sampling with class distribution balancing for active annotation in aerial object detection. IEEE Trans. Geosci. Remote Sens. 61, 1–13. https://doi.org/10.1109/TGRS.2023.3285443.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 2117–2125. https://doi.org/10.1109/CVPR.2017.106.

Lin, Y., Feng, P., Guan, J., Wang, W., Chambers, J., 2019. IENet: Interacting Embranchment One Stage Anchor Free Detector for Orientation Aerial Object Detection. https://doi.org/10.48550/arXiv.1912.00969.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017b. Focal loss for dense object detection. In: Proc. IEEE Int. Conf. Comput. vis., pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.324.

Lin, Y., He, H., Yin, Z., Chen, F., 2014. Rotation-invariant object detection in remote sensing images based on radial-gradient angle. IEEE Geosci. Remote. Sens. Lett. 12 (4), 746–750. https://doi.org/10.1109/LGRS.2014.2360887.

Liu, H., Hu, Y., 2022. Relationship Reasoning with Triple Attention Network (RR-TAN) for object detection of remote sensing images. In: Proc. IEEE Int. Geosci. Remote Sens. Symp., pp. 2670–2673. https://doi.org/10.1109/IGARSS46834.2022.9884662.

Liu, K., Mattyus, G., 2015. Fast multiclass vehicle detection on aerial images. IEEE Geosci. Remote. Sens. Lett. 12, 1938–1942. https://doi.org/10.1109/LGRS.2015.2439517.

Liu, G., Sun, X., Fu, K., Wang, H., 2012. Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. IEEE Geosci. Remote. Sens. Lett. 10 (3), 573–577. https://doi.org/10.1109/LGRS.2012.2214022.

Liu, Z., Wang, H., Weng, L., Yang, Y., 2016. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. IEEE Geosci. Remote Sens. Lett. 13, 1074–1078. https://doi.org/10.1109/LGRS.2016.2565705.

Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 2486–2498. https://doi.org/10.1109/TGRS.2016.2645610.

Lu, D., Li, D., Li, Y., Wang, S., 2022. OSKDet: orientation-sensitive keypoint localization for rotated object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 1172–1182. https://doi.org/10.1109/CVPR52688.2022.00125.

Lyu, C., Zhang, W., Huang, H., Zhou, Y., Wang, Y., Liu, Y., Zhang, S., Chen, K., 2022. Rtmdet: An empirical study of designing real-time object detectors. https://doi.org/10.48550/arXiv.2212.07784.

Ma, T., Mao, M., Zheng, H., Gao, P., Wang, X., Han, S., Ding, E., Zhang, B., Doermann, D., 2021. Oriented Object Detection with Transformer. https://doi.org/10.48550/arXiv.2106.03146.

Milz, S., Rüdiger, T., Süss, S., 2018. Aerial ganeration towards realistic data augmentation using conditional gans. In: Proc. Eur. Conf. Comput. vis. Workshops., pp. 59–72. https://doi.org/10.1007/978-3-030-11012-3_5.

Ming, Q., Zhou, Z., Miao, L., Zhang, H., Li, L., 2020. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. In: Proc. AAAI Conf. Artif. Intell. vol. 35(3), pp. 2355-2363. https://doi.org/10.1609/aaai.v35i3.16336.

Ming, Q., Miao, L., Zhou, Z., Dong, Y., 2022a. CFC-Net: a critical feature capturing network for arbitrary-oriented object detection in remote-sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3095186.

Ming, Q., Miao, L., Zhou, Z., Yang, X., Dong, Y., 2022b. Optimization for arbitrary-oriented object detection via representation invariance loss. IEEE Geosci. Remote. Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3115110.

Moon, H., Chellappa, R., Rosenfeld, A., 2002. Performance analysis of a simple vehicle detection algorithm. Image. vis. Comput. 20 (1), 1–13. https://doi.org/10.1016/S0262-8856(01)00059-2.

Mundhenk, T.N., Konjevod, G., Sakla, W.A., Boakye, K., 2016. A large contextual dataset for classification, detection and counting of cars with deep learning. In: Proc. Eur. Conf. Comput. vis., pp. 785–800. https://doi.org/10.1007/978-3-319-46487-9_48.

Murrugarra-Llerena, J., Zeni, L.F., Kristen, L.N., Jung, C., 2021. Gaussian Bounding Boxes and Probabilistic Intersection-over-Union for Object Detection. https://doi.org/10.48550/arXiv.2106.06072.

Pan, X., Ren, Y., Sheng, K., Dong, W., Yuan, H., Guo, X., Ma, C., Xu, C., 2020. Dynamic refinement network for oriented and densely packed object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 11204–11213. https://doi.org/10.1109/cvpr42600.2020.01122.

Pu, Y., Wang, Y., Xia, Z., Han, Y., Wang, Y., Gan, W., Wang, Z., Song, S., Huang, G., 2023. Adaptive rotated convolution for rotated object detection. In: Proc. IEEE Int. Conf. Comput. vis., pp. 6566–6577. https://doi.org/10.1109/ICCV51070.2023.00606.

Qian, W., Yang, X., Peng, S., Guo, Y., Yan, J., 2019. Learning Modulated Loss for Rotated Object Detection. In: Proc. AAAI Conf. Artif. Intell. vol. 35(3), pp. 2458-2466. https://doi.org/10.1609/aaai.v35i3.16347.

Qian, W., Yang, X., Peng, S., Zhang, X., Yan, J., 2022. RSDet++: Point-based modulated loss for more accurate rotated object detection. IEEE Trans. Circuits Syst. Video Technol. 32 (11), 7869–7879. https://doi.org/10.1109/TCSVT.2022.3186070.

Qin, R., Liu, Q., Gao, G., Huang, D., Wang, Y., 2020. MRDet: a multi-head network for accurate oriented object detection in aerial images. IEEE Trans. Geosci. Remote Sens. 60, 1–12. https://doi.org/10.1109/TGRS.2021.3113473.

Qu, Z., Du, J., Cao, Y., Guan, Q., Zhao, P., 2020. Deep Active Learning for Remote Sensing Object Detection. https://doi.org/10.48550/arXiv.2003.08793.

Razakarivony, S., Jurie, F., 2016. Vehicle detection in aerial imagery: A small target detection benchmark. J. vis. Commun. Image Represent. 34, 187–203. https://doi.org/10.1016/j.jvcir.2015.11.002.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Fei-Fei, L., 2014. ImageNet large scale visual recognition challenge. Int. J. Comput. vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sairam, R.V.C., Keswani, M., Sinha, U., Shah, N., Balasubramanian, V.N., 2023. ARUBA: An Architecture-Agnostic Balanced Loss for Aerial Object Detection. In: Proc. IEEE Winter Conf. Appl. Comput. Vis. pp. 3708-3717. https://doi.org/10.1109/WACV56688.2023.00371.

Shamsolmoali, P., Zareapoor, M., Chanussot, J., Zhou, H., Yang, J., 2021. Rotation equivariant feature image pyramid network for object detection in optical remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3112481.

Shamsolmoali, P., Chanussot, J., Zareapoor, M., Zhou, H., Yang, J., 2022a. Multipatch feature pyramid network for weakly supervised object detection in optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2021.3106442.

Shamsolmoali, P., Zareapoor, M., Yang, J., Granger, E., Chanussot, J., 2022b. Enhanced single-shot detector for small object detection in remote sensing images. In: Proc. IEEE Int. Geosci. Remote Sens. Symp., pp. 1716–1719. https://doi.org/10.1109/IGARSS46834.2022.9884546.

Shen, Y., Zhang, D., Song, Z., Jiang, X., Ye, Q., 2023. Learning to reduce information bottleneck for object detection in aerial images. IEEE Geosci. Remote. Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3264455.

Shermeyer, J., Hossler, T., Van Etten, A., Hogan, D., Lewis, R., Kim, D., 2021. Rareplanes: Synthetic data takes flight. In: Proc. IEEE Winter Conf. Appl. Comput. Vis. pp. 207-217. https://doi.org/10.1109/WACV48630.2021.00025.

Shermeyer, J., Van Etten, A., 2019. The effects of super-resolution on object detection performance in satellite imagery. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog. Workshops., pp. 1432–1441. https://doi.org/10.1109/CVPRW.2019.00184.

Stumpf, A., Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. Remote Sens. Environ. 115 (10), 2564–2577. https://doi.org/10.1016/j.rse.2011.05.013.

Sun, P., Chen, G., Luke, G., Shang, Y., 2018. Salience biased loss for object detection in aerial images. https://doi.org/10.48550/arXiv.1810.08103.

Sun, X., Wang, P., Yan, Z., Xu, F., Wang, R., Diao, W., Chen, J., Li, J., Feng, Y., Xu, T., Weinmann, M., 2022. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 184, 116–130. https://doi.org/10.1016/j.isprsjprs.2021.12.004.

Tang, J., Zhang, W., Liu, H., Yang, M., Jiang, B., Hu, G., Bai, X., 2022. Few could be better than all: feature sampling and grouping for scene text detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 4553–4562. https://doi.org/10.1109/CVPR52688.2022.00452.

Tao, C., Tan, Y., Cai, H., Tian, J., 2010. Airport detection from large IKONOS images using clustered SIFT keypoints and region information. IEEE Geosci. Remote. Sens. Lett. 8 (1), 128–132. https://doi.org/10.1109/LGRS.2010.2051792.

Torralba, A., Efros, A.A., 2011. Unbiased look at dataset bias. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 1521–1528. https://doi.org/10.1109/CVPR.2011.5995347.

Varga, L.A., Zell, A., 2021. Tackling the background bias in sparse object detection via cropped windows. In: Proc. IEEE Int. Conf. Comput. vis. Workshops., pp. 2768–2777. https://doi.org/10.1109/ICCVW54120.2021.00311.

Wang, X., Cai, Z., Gao, D., Vasconcelos, N., 2019. Towards universal object detection by domain attention. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 7289–7298. https://doi.org/10.1109/CVPR.2019.00746.

Wang, X., Wang, G., Dang, Q., Liu, Y., Hu, X., Yu, D., 2022b. PP-YOLOE-R: An Efficient Anchor-Free Rotated Object Detector. https://doi.org/10.48550/arXiv.2211.02386.

Wang, H., Huang, Z., Chen, Z., Song, Y., Li, W., 2022a. Multigrained angle representation for remote-sensing object detection. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2022.3212592.

Wang, F., Newkirk, R., 1988. A knowledge-based system for highway network extraction. IEEE Trans. Geosci. Remote Sens. 26 (5), 525–531. https://doi.org/10.1109/36.7677.

Wang, J., Yang, W., Li, H.-C., Zhang, H., Xia, G.-S., 2020. Learning center probability map for detecting objects in aerial images. IEEE Trans. Geosci. Remote Sens. 59, 4307–4323. https://doi.org/10.1109/TGRS.2020.3010051.

Wang, J., Yang, W., Guo, H., Zhang, R., Xia, G.-S., 2021. Tiny object detection in aerial images. In: Proc. Int. Conf. Pattern Recog., pp. 3791–3798. https://doi.org/10.1109/ICPR48806.2021.9413340.

Wang, D., Zhang, J., Du, B., Xia, G.-S., Tao, D., 2023a. An empirical study of remote sensing pretraining. IEEE Trans. Geosci. Remote Sens. 61, 1–20. https://doi.org/10.1109/TGRS.2022.3176603.

Wang, D., Zhang, Q., Xu, Y., Zhang, J., Du, B., Tao, D., Zhang, L., 2023b. Advancing plain vision transformer toward remote sensing foundation model. IEEE Trans. Geosci. Remote Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2022.3222818.

Wang, D., Zhang, J., Xu, M., Liu, L., Wang, D., Gao, E., Han, C., Guo, H., Du, B., Tao, D., Zhang, L., 2024. MTP: Advancing remote sensing foundation model via multi-task pretraining. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 17, 11632–11654. https://doi.org/10.1109/JSTARS.2024.3408154.

Waqas Zamir, S., Arora, A., Gupta, A., Khan, S., Sun, G., Shahbaz Khan, F., Zhu, F., Shao, L., Xia, G.-S., Bai, X., 2019. isaid: A large-scale dataset for instance

segmentation in aerial images. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog. Workshops., pp. 28–37. https://doi.org/10.48550/arXiv.1905.12886.
Weber, J., Lefevre, S., 2008. A multivariate hit-or-miss transform for conjoint spatial and spectral template matching. In: Proc. IEEE Int. Conf. Image Signal Process. pp. 226-235. https://doi.org/10.1007/978-3-540-69905-7_26.
Weber, J., Lefèvre, S., 2012. Spatial and spectral morphological template matching. Image. vis. Comput. 30 (12), 934–945. https://doi.org/10.1016/j.imavis.2012.07.002.
Wei, H., Zhou, L., Zhang, Y., Li, H., Guo, R., Wang, H., 2019. Oriented objects as pairs of middle lines. ISPRS J. Photogramm. Remote Sens. 169, 268–279. https://doi.org/10.1016/j.isprsjprs.2020.09.022.
Weir, N., Lindenbaum, D., Bastidas, A., Etten, A.V., McPherson, S., Shermeyer, J., Kumar, V., Tang, H., 2019. Spacenet mvoi: A multi-view overhead imagery dataset. In: Proc. IEEE Int. Conf. Comput. vis., pp. 992–1001. https://doi.org/10.1109/ICCV.2019.00108.
Wen, S., Guo, W., Liu, Y., Wu, R., 2022. Rotated object detection via scale-invariant mahalanobis distance in aerial images. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2022.3197617.
Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A large-scale dataset for object detection in aerial images. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 3974–3983. https://doi.org/10.1109/CVPR.2018.00418.
Xiao, J., Yao, Y., Zhou, J., Guo, H., Yu, Q., Wang, Y.-F., 2023. FDLR-Net: A feature decoupling and localization refinement network for object detection in remote sensing images. Expert Syst. Appl. 225, 120068. https://doi.org/10.1016/j.eswa.2023.120068.
Xie, X., Cheng, G., Wang, J., Yao, X., Han, J., 2021. Oriented R-CNN for Object Detection. In: Proc. IEEE Int. Conf. Comput. vis., pp. 3500–3509. https://doi.org/10.1109/ICCV48922.2021.00350.
Xu, C., Wang, J., Yang, W., Yu, L., 2021a. Dot distance for tiny object detection in aerial images. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog. Workshops., pp. 1192–1201. https://doi.org/10.1109/CVPRW53098.2021.00130.
Xu, C., Ding, J., Wang, J., Yang, W., Yu, H., Yu, L., Xia, G.-S., 2023. Dynamic coarse-to-fine learning for oriented tiny object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 7318–7328. https://doi.org/10.1109/CVPR52729.2023.00707.
Xu, J., Li, Y., Wang, S., 2021b. AdaZoom: Adaptive Zoom Network for Multi-Scale Object Detection in Large Scenes. https://doi.org/10.48550/arXiv.2106.10409.
Xu, Y., Fu, M., Wang, Q., Wang, Y., Chen, K., Xia, G.S., Bai, X., 2021c. Gliding vertex on the horizontal bounding box for multi-oriented object detection. IEEE Trans. Pattern Anal. Mach. Intell. 43, 1452–1459. https://doi.org/10.1109/TPAMI.2020.2974745.
Yang, X., Yan, J., Feng, Z., He, T., 2019b. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. In: Proc. AAAI Conf. Artif. Intell. vol. 35(4), pp. 3163-3171. https://doi.org/10.1609/aaai.v35i4.16426.
Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X., Tian, Q., 2021d. Rethinking rotated object detection with gaussian wasserstein distance loss. In: Proc. Int. Conf. Mach. Learn. Appl. pp. 11830-11841. https://doi.org/10.48550/arXiv.2101.11952.
Yang, Y., Chen, J., Zhong, X., Deng, Y., 2022c. Polygon-to-polygon distance loss for rotated object detection. In: Proc. AAAI Conf. Artif. Intell. vol. 36(3), pp. 3072-3080. https://doi.org/10.1609/aaai.v36i3.20214.
Yang, X., Zhang, G., Li, W., Wang, X., Zhou, Y., Yan, J., 2023. H2RBox: Horizontal Box Annotation is All You Need for Oriented Object Detection. In: Proc. Int. Conf. Learn. Represent. https://doi.org/10.48550/arXiv.2210.06742.
Yang, X., Zhou, Y., Zhang, G., Yang, J., Wang, W., Yan, J., Zhang, X., Tian, Q., 2023. The KFIoU Loss for Rotated Object Detection. In: Proc. Int. Conf. Learn. Represent. https://doi.org/10.48550/arXiv.2201.12558.
Yang, F., Fan, H., Chu, P., Blasch, E., Ling, H., 2019a. Clustered object detection in aerial images. In: Proc. IEEE Int. Conf. Comput. vis., pp. 8310–8319. https://doi.org/10.1109/ICCV.2019.00840.
Yang, X., Hou, L., Zhou, Y., Wang, W., Yan, J., 2021b. Dense label encoding for boundary discontinuity free rotation detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 15814–15824. https://doi.org/10.1109/CVPR46437.2021.01556.
Yang, G.-Y., Li, X.-L., Xiao, Z.-K., Mu, T.-J., Martin, R.R., Hu, S.-M., 2021a. Sampling equivariant self-attention networks for object detection in aerial images. IEEE Trans. Image Process. 32, 6413–6425. https://doi.org/10.1109/TIP.2023.3327586.
Yang, M.Y., Liao, W., Li, X., Rosenhahn, B., 2018. Deep learning for vehicle detection in aerial images. In: Proc. IEEE Int. Conf. Inf. Process., pp. 3079–3083. https://doi.org/10.1109/ICIP.2018.8451454.
Yang, X., Yan, J., 2020. Arbitrary-oriented object detection with circular smooth label. In: Proc. Eur. Conf. Comput. vis., pp. 677–694. https://doi.org/10.1007/978-3-030-58598-3_40.
Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., Sun, X., Fu, K., 2019c. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In: Proc. IEEE Int. Conf. Comput. vis., pp. 8232–8241. https://doi.org/10.1109/ICCV.2019.00832.
Yang, X., Yan, J., He, T., 2020a. On the arbitrary-oriented object detection: classification based approaches revisited. Int. J. Comput. vis. 130, 1340–1365. https://doi.org/10.1007/s11263-022-01593-w.
Yang, X., Yan, J., Yang, X., Tang, J., Liao, W., He, T., 2020b. SCRDet++: detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. IEEE Trans. Pattern Anal. Mach. Intell. 45, 2384–2399. https://doi.org/10.1109/TPAMI.2022.3166956.
Yang, X., Yang, X., Yang, J., Ming, Q., Wang, W., Tian, Q., Yan, J., 2021e. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence. In: Proc. Adv. Neural Inf. Process. Syst., pp. 18381–18394. https://doi.org/10.48550/arXiv.2106.01883.
Yi, J., Wu, P., Liu, B., Huang, Q., Qu, H., Metaxas, D., 2021. Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. In: Proc. IEEE Winter Conf. Appl. Comput. Vis. pp. 2149-2158. https://doi.org/10.1109/WACV48630.2021.00220.
Yu, Y., Da, F., 2023. Phase-shifting coder: predicting accurate orientation in oriented object detection. In: Proc. IEEE Conf. Comput. vis. Pattern. Recog., pp. 13354–13363. https://doi.org/10.1109/CVPR52729.2023.01283.
Yu, H., Tian, Y., Ye, Q., Liu, Y., 2024, March. Spatial transform decoupling for oriented object detection. In: Proc. AAAI Conf. Artif. Intell. vol. 38(7), pp. 6782-6790. https://doi.org/10.1609/aaai.v38i7.28502.
Yu, Y., Yang, X., Li, Q., Zhou, Y., Zhang, G., Da, F., Yan, J., 2023. H2RBox-v2: Boosting HBox-supervised oriented object detection via symmetric learning. In: Proc. Adv. Neural Inf. Process. Syst.. https://doi.org/10.48550/arXiv.2304.04403.
Zand, M., Etemad, A., Greenspan, M., 2022. Oriented bounding boxes for small and freely rotated objects. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2021.3076050.
Zeng, Y., Yang, X., Li, Q., Chen, Y., Yan, J., 2023. ARS-DETR: aspect ratio sensitive oriented object detection with transformer. IEEE Trans. Geosci. Remote Sens. 62, 1–15. https://doi.org/10.1109/TGRS.2024.3364713.
Zhang, D., Han, J., Cheng, G., Yang, M.-H., 2021. Weakly supervised object localization and detection: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44, 5866–5885. https://doi.org/10.1109/TPAMI.2021.3074313.
Zhang, J., Lei, J., Xie, W., Fang, Z., Li, Y., Du, Q., 2023. SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3258666.
Zhang, G., Lu, S., Zhang, W., 2019a. CAD-Net: a context-aware detection network for objects in remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 57, 10015–10024. https://doi.org/10.1109/TGRS.2019.2930982.
Zhang, Z., Vosselman, G., Gerke, M., Persello, C., Tuia, D., Yang, M.Y., 2019c. Detecting building changes between airborne laser scanning and photogrammetric data. Remote Sens. 11, 2417. https://doi.org/10.3390/rs11202417.
Zhang, F., Wang, X., Zhou, S., Wang, Y., 2022a. DARDet: a dense anchor-free rotated object detector in aerial images. IEEE Geosci. Remote. Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3122924.
Zhang, Y., Yuan, Y., Feng, Y., Lu, X., 2019b. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. IEEE Trans. Geosci. Remote Sens. 57, 5535–5548. https://doi.org/10.1109/TGRS.2019.2900302.
Zhang, T., Zhuang, Y., Wang, G., Dong, S., Chen, H., Li, L., 2022b. Multiscale semantic fusion-guided fractal convolutional object detection network for optical remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 60, 1–20. https://doi.org/10.1109/tgrs.2021.3108476.
Zhao, P., Qu, Z., Bu, Y., Tan, W., Guan, Q., 2021. Polardet: A fast, more precise detector for rotated target in aerial images. Int. J. Remote. Sens. 42, 5831–5861. https://doi.org/10.1080/01431161.2021.1931535.
Zheng, Z., Ye, R., Hou, Q., Ren, D., Wang, P., Zuo, W., Cheng, M.M., 2023. Localization distillation for object detection. IEEE Trans. Pattern Anal. Mach. Intell. 45, 10070–10083. https://doi.org/10.1109/TPAMI.2023.3248583.
Zhou, Y., Yang, X., Zhang, G., Wang, J., Liu, Y., Hou, L., Jiang, X., Liu, X., Yan, J., Lyu, C., Zhang, W., Chen, K., 2022. MMRotate: A Rotated Object Detection Benchmark using PyTorch. In: Proc. ACM Trans. Multimedia Comput. Commun. Appl. pp. 7331-7334. https://doi.org/10.1145/3503161.3548541.
Zhou, Q., Yu, C., Wang, Z., Wang, F., 2023. D2Q-DETR: Decoupling and Dynamic Queries for Oriented Object Detection with Transformers. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. pp. 1-5. https://doi.org/10.1109/ICASSP49357.2023.10095341.
Zhu, H., Chen, X., Dai, W., Fu, K., Ye, Q., Jiao, J., 2015. Orientation robust object detection in aerial images using deep convolutional neural network. In: Proc. IEEE Int. Conf. Inf. Process., pp. 3735–3739. https://doi.org/10.1109/ICIP.2015.7351502.
Zhu, Y., Du, J., Wu, X., 2020. Adaptive period embedding for representing oriented objects in aerial images. IEEE Trans. Geosci. Remote Sens. 58, 7247–7257. https://doi.org/10.1109/TGRS.2020.2981203.
Zhu, P., Wen, L., Bian, X., Ling, H., Hu, Q., 2018. Vision meets drones: A challenge. https://doi.org/10.48550/arXiv.1804.07437.
Zou, Z., Shi, Z., 2017. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images. IEEE Trans. Image Process. 27, 1100–1111. https://doi.org/10.1109/TIP.2017.2773199.

Further reading

Lee, C., Son, J., Shon, H., Jeon, Y., Kim, J., 2024. FRED: Towards a Full Rotation-Equivariance in Aerial Image Object Detection. In: Proc. AAAI Conf. Artif. Intell. vol. 38(4), pp. 2883-2891. https://doi.org/10.48550/arXiv.2401.06159.
Li, Z., Hou, B., Wu, Z., Ren, B., Yang, C., 2023b. FCOSR: a simple anchor-free rotated detector for aerial object detection. Remote Sens. 15, 5499. https://doi.org/10.3390/rs15235499.
Wang, C., Guo, G., Liu, C., Shao, D., Gao, S., 2024b. Effective rotate: learning rotation-robust prototype for aerial object detection. IEEE Trans. Geosci. Remote Sens. 62, 1–14. https://doi.org/10.1109/TGRS.2024.3374880.
Wang, D., Zhang, J., Du, B., Xu, M., Liu, L., Tao, D., Zhang, L., 2024b. Samrs: Scaling-up remote sensing segmentation dataset with segment anything model. In: Proc. Adv. Neural Inf. Process. Syst.. https://doi.org/10.48550/arXiv.2305.02034.