remote sensing

MDPI

Article

KFGOD: A Fine-Grained Object Detection Dataset in KOMPSAT Satellite Imagery

Dong Ho Lee [1], Ji Hun Hong [1], Hyun Woo Seo [2] and Han Oh [1,2,*]

[1] Korea Aerospace Research Institute (KARI), 169-84 Gwahak-ro, Yuseong-gu, Daejeon 34133, Republic of Korea; ehdgh3337@kari.re.kr (D.H.L.); jhhong@kari.re.kr (J.H.H.)
[2] Department of Aerospace System Engineering, University of Science and Technology (UST), 169-84 Gwahak-ro, Yuseong-gu, Daejeon 34133, Republic of Korea; shw0106@kari.re.kr
* Correspondence: ohhan@kari.re.kr

Highlights

What are the main findings?

- KFGOD provides approximately 880K object instances across 33 fine-grained classes from homogeneous KOMPSAT-3/3A imagery (0.55–0.7 m resolution), with dual OBB+HBB annotations.
- The dataset's unique sensor homogeneity (KOMPSAT-3/3A only) provides a well-controlled, sensor-consistent benchmark that minimizes sensor-induced domain gaps and enables a fair comparison of the detection algorithms.

What are the implications of the main findings?

- Benchmark results (SOTA mAP 63.9%) validate KFGOD as a challenging benchmark, highlighting critical, real-world research problems in fine-grained, long-tail, and oriented object detection.
- Multi-format label support and demonstrated real-world use cases (e.g., Korea Coast Guard maritime surveillance) show that KFGOD is practically useful and generalizes well to diverse high-resolution satellite imagery.

Abstract

Object detection in high-resolution satellite imagery is a critical technology for various applications, yet it faces persistent challenges due to extreme variations in object scale, orientation, and density. The development of numerous public datasets has been pivotal for advancing the field. To continue this progress and expand the diversity of sensor data available for research, we introduce the KOMPSAT Fine-Grained Object Detection (KFGOD) dataset, a new large-scale benchmark for fine-grained object detection. KFGOD is uniquely constructed using 70 cm and 55 cm resolution optical imagery from the KOMPSAT-3 and 3A satellites, sources not covered by existing major datasets. It provides approximately 880,000 object instances across 33 fine-grained classes, encompassing a wide range of ships, aircraft, vehicles, and infrastructure. The dataset ensures high quality and sensor consistency, covering diverse geographical regions worldwide to promote model generalization. For precise localization, all objects are annotated with both oriented (OBB) and horizontal (HBB) bounding boxes. Comprehensive experiments with state-of-the-art detection models provide benchmark results and highlight the challenging nature of the dataset, particularly in distinguishing between visually similar fine-grained classes. The KFGOD dataset is publicly available and aims to foster further research in fine-grained object detection and analysis of high-resolution satellite imagery.

Academic Editors: Yanni Dong, Xiaochen Yang and Qian Du

Received: 6 October 2025
Revised: 8 November 2025
Accepted: 18 November 2025
Published: 20 November 2025

Citation: Lee, D.H.; Hong, J.H.; Seo, H.W.; Oh, H. KFGOD: A Fine-Grained Object Detection Dataset in KOMPSAT Satellite Imagery. *Remote Sens.* **2025**, *17*, 3774. https://doi.org/10.3390/rs17223774

Keywords: fine-grained object detection; KOMPSAT; large-scale dataset; satellite imagery; artificial intelligence; remote sensing

1. Introduction

Recently, the field of Earth observation (EO) has been rapidly advancing, driven by improvements in satellite sensors and data processing capabilities. In particular, the increased availability of high-resolution satellite imagery has underscored the importance of object detection techniques for automatically identifying and classifying specific objects within images across a broad spectrum of applications, including national defense and security, urban planning, resource management, environmental monitoring, and disaster response [1–8]. Object detection in satellite imagery plays a pivotal role in periodically monitoring vast areas and analyzing changes, thereby providing foundational information for rapid and accurate decision-making [9,10]. While deep learning advancements have achieved remarkable performance in object detection for natural images [11–16], satellite images exhibit fundamental differences from them in terms of imaging environments and object characteristics, which limits the direct application of existing techniques [17,18].

Satellite images are typically captured from a top–down perspective, overlooking the Earth's surface vertically or near vertically. This imaging approach results in the loss of three-dimensional shape information for objects, and even objects of the same class can appear significantly different in the projected image depending on their ground-level orientation, leading to pronounced rotational variance [19,20]. Furthermore, a single image may contain objects exhibiting extreme scale variations, ranging from small vehicles spanning a few pixels to large ships or buildings extending across hundreds or thousands of pixels, which intensifies the challenge for detection models to robustly handle diverse scales [21,22]. Additionally, objects of the same category can vary markedly in appearance depending on the season or region of capture. Moreover, in certain areas such as ports or urban centers, objects are densely clustered, while in other expansive regions, they are sparsely distributed, resulting in severe density imbalances [23–26]. These complex and inherent characteristics of satellite imagery pose major technical challenges that degrade the performance of general object detection algorithms, necessitating specialized approaches tailored to satellite images [2,4,18].

In response to these challenges, numerous public datasets for object detection in satellite and aerial imagery, such as large-scale Dataset of Object deTection in Aerial images (DOTA), xView, Fine-grAined object recognition in high-Resolution remote sensing imagery (FAIR1M), and Object Detection in Optical Remote sensing images (DIOR), have been released in recent years. These datasets have provided a common foundation for algorithm development and performance evaluation, significantly advancing related technologies [4,9,10,27–29]. However, many prominent large-scale benchmarks, while valuable for testing generalization, are inherently heterogeneous, often combining imagery from various sensors with different GSDs (Ground Sample Distances) and processing levels. This heterogeneity can introduce confounding variables, such as the domain gaps between sensors, making it challenging to isolate and fairly evaluate the core performance of detection algorithms themselves. Consequently, a notable gap exists for a large-scale, high-resolution benchmark with high data homogeneity designed to facilitate fair and focused algorithm comparison. To address this specific gap, this study introduces the KOMPSAT Fine-Grained Object Detection (KFGOD) dataset, built exclusively from the KOMPSAT-3 and 3A satellite series.

The KFGOD dataset is proposed to address this gap, and offers the following distinctive contributions.

First, and most critically, the dataset provides a homogeneous benchmark for fair algorithm comparison. The dataset's primary contribution is its intentional design for data homogeneity. Unlike heterogeneous benchmarks that source images from multiple sensors, KFGOD is constructed exclusively from the KOMPSAT-3/3A single sensor family. This approach ensures a highly consistent spatial resolution (0.55–0.7 m) and uniform sensor characteristics across the entire dataset. By intentionally controlling for the confounding variable of sensor-based domain gaps, KFGOD provides a stable foundation for researchers to train, test, and fairly compare the core performance of object detection models, isolating algorithmic improvements from domain adaptation effects.

Second, it encompasses broad geographical diversity to promote model generalization. Designed to include various cities and terrains across multiple continents, the dataset enables models trained on it to achieve robust generalization across diverse geographical environments. Based on this, 33 classes are defined, and oriented bounding boxes (OBBs) that precisely represent the rotation angles of objects are provided alongside standard horizontal bounding boxes (HBBs). OBB annotations clarify the orientation of elongated objects such as ships or aircraft and minimize the inclusion of unnecessary background information, enabling more precise localization.

Third, it is designed for broad usability and accessibility. For user convenience, the dataset provides labels in multiple formats, including COCO, DOTA, GeoJSON, and YOLO, which can be directly used without additional conversion in the most widely adopted object detection frameworks and geographic information systems (GISs).

The structure of this paper is as follows. Section 2 summarizes the key existing satellite imagery object detection datasets and derives the necessity and distinctiveness of the proposed dataset based on this summary. Section 3 describes the construction process of the KFGOD dataset based on KOMPSAT-3/3A satellite images, detailing image collection, preprocessing, annotation strategies, and class definition methods. Section 4 analyzes various statistical characteristics of the dataset, such as class distributions, instance sizes, and image densities, to illuminate its structural features. Section 5 validates the effectiveness of the dataset by comparing and evaluating the performance of state-of-the-art object detection models using KFGOD. Finally, Section 6 summarizes the conclusions of this study and discusses future utilization and expansion possibilities for the dataset.

2. Related Work

The field of object detection in satellite imagery has seen significant progress in recent years, driven by enhancements in high-resolution satellite sensors and deep learning methodologies [1–3,18]. Early research primarily revolved around small-scale datasets such as NWPU VHR-10 [30] and VEDAI [27]. However, the emergence of large-scale benchmarks like xView [4], DOTA [9], DIOR [3], and FAIR1M [10] has led to substantial progress in terms of object counts, class diversity, spatial resolution, and annotation precision.

The xView dataset, based on Maxar's WorldView-3 satellite images, is a representative large-scale dataset that provides over one million object instances and 60 broad classes in an ultra-high-resolution environment of 30 cm/pixel [4]. This vast scale and diversity have significantly contributed to the advancement of fine-grained object recognition research, with annotations provided in the form of HBBs that do not incorporate object orientation information.

DOTA serves as a leading benchmark for rotated object detection by offering OBB annotations. It leverages images from Google Earth and multiple satellite sources to create an environment suitable for learning and evaluating objects of various shapes and scales [9].

Comprising images from diverse sensors, it encompasses a wide range of resolutions from 0.1 m/pixel to 4.5 m/pixel, which offers advantages in testing model robustness. However, the annotations do not include geographic information.

DIOR consists of 23,463 satellite images, covering various countries, seasons, and resolutions (0.5–30 m/pixel), with HBB annotations provided for 20 classes [3].

FAIR1M is a prominent benchmark specialized in fine-grained recognition, enabling detailed differentiation among similar objects [10]. Drawing from over 40,000 ultra-high-resolution (0.3–0.8 m/pixel) images sourced from Gaofen and Google Earth satellites, this dataset offers more than one million object instances across 5 major categories and 37 sub-classes. A distinctive feature is that all object instances are annotated with OBBs, which precisely capture the orientation of rotated objects.

Despite significant advancements in object counts and class diversity, several challenges persist, for instance, biases toward certain classes or insufficient samples for rare objects, difficulties in evaluating detection performance for small objects or high-density scenes, limitations of HBB-based annotations, and the inherent heterogeneity of benchmarks that combine multiple sensors. While valuable for testing generalization, this mixing of sensors, resolutions, and processing levels introduces domain gaps that act as confounding variables, making it challenging to fairly assess the core performance of algorithms themselves [17–19,21].

In light of these issues, recent studies have emphasized the need for task-specific datasets tailored to applications such as traffic infrastructure monitoring [21,31,32], disaster response [33], and defense surveillance [22]. In particular, there is growing demand for object subcategorization, extensive regional coverage, robustness to small objects and dense scenes, accurate orientation annotations, and the provision of refined geographic information [2].

To address these requirements, this study proposes a KOMPSAT-based object detection dataset. The high-resolution images acquired from the KOMPSAT-3/3A satellites span various regions and periods over a decade, ensuring excellent data homogeneity—specifically, a highly consistent resolution and set of sensor characteristics—due to their origin from the same sensor series. This homogeneous approach directly addresses the challenge of confounding variables present in heterogeneous datasets, enabling a fairer comparison of algorithmic performance. For a total of 33 detailed classes, both OBB and HBB annotations are provided in formats such as COCO (HBB), DOTA (HBB, OBB), and YOLO (HBB, OBB), with support for GeoJSON to facilitate use in diverse detection models and GIS analyses. All object instances incorporate geographic information through precise positional alignment, and the dataset reflects variations in object density and size to enable balanced evaluation of detection performance in realistic complex environments (Table 1).

Table 1. Summary of public remote sensing object detection datasets.

Dataset	Source	Instances [a]	Images	Image Width (px)	Categories	Annotation	Format	Fine-Grained [b]
NWPU VHR-10 [30]	Google Earth	3775	800	~1000	10	HBB	JPG	N
VEDAI [27]	Google Earth	3640	1210	512, 1024	9	OBB	PNG	Y
UCAS-AOD [34]	Google Earth	6029	910	~1000	2	OBB	PNG	N
HRSC2016 [24]	Google Earth	2976	1070	~1100	1	OBB	BMP	N
DOTA [9]	Google Earth, Satellite JL-1, GF-2	188,282	2806	800–4000	15	OBB + HBB	PNG	N
HRRSD [35]	Google Earth, Baidu Map	55,740	21,761	152–10,569	13	HBB	JPG	N

Table 1. *Cont.*

Dataset	Source	Instances [a]	Images	Image Width (px)	Categories	Annotation	Format	Fine-Grained [b]
RSOD [36]	Google Earth, Tianditu	6950	976	∼1000	4	HBB	JPG	N
xView [4]	WorldView-3	1 M	1127	2000–4000	60	HBB	PNG	Y
DIOR [3]	Google Earth	192,472	23,463	800	20	HBB	JPG	N
FGSD [29]	Google Earth	5634	2612	930	43	OBB	JPG	Y
FAIR1M [10]	Gaofen, Google Earth	1.02 M	42,796	600–10,000	37	OBB	TIFF	Y
KFGOD (Ours)	KOMPSAT-3, KOMPSAT-3A	882,399	4003	1024	33	OBB + HBB	PNG	Y

[a] M: Million (1,000,000). [b] Y: Yes, N: No.

3. Dataset Construction

This section details the construction process of the KFGOD dataset, an object detection dataset built upon high-resolution optical satellite imagery collected from the KOMPSAT-3/3A satellites. The KFGOD dataset was developed through a systematic series of procedures, including high-quality image acquisition, preprocessing, labeling (both OBB and HBB), hierarchical class definition, and multi-stage quality assurance. This section provides a comprehensive description of the methods employed in constructing the KFGOD dataset.

3.1. Image Acquisition and Preprocessing

3.1.1. KOMPSAT-3/3A Data Collection

KOMPSAT-3/3A are high-resolution Earth observation satellites developed by the Korea Aerospace Research Institute (KARI), launched in 2012 and 2015, respectively, and have been operating stably to date. KOMPSAT-3 provides optical imagery with a spatial resolution of 70 cm in nadir view from an orbital altitude of 685 km, while KOMPSAT-3A can acquire precise imagery at 55 cm resolution from an altitude of 528 km (Table 2).

Table 2. Technical specifications of KOMPSAT-3 and KOMPSAT-3A satellites.

Characteristic	KOMPSAT-3	KOMPSAT-3A
Sensor	Optical	
Orbital Altitude	685 km	528 km
Spatial Resolution	Pan: 70 cm, MS: 4 m	Pan: 55 cm, MS: 3.2 m
Band Configuration	Panchromatic, Blue, Green, Red, NIR	Panchromatic, Blue, Green, Red, NIR, IR
Image Size	24,000 × 24,000 (px)	
Orbit Type	Sun-Synchronous	

In this study, a dataset for training object detection models was constructed based on optical satellite imagery collected from the KOMPSAT-3 and KOMPSAT-3A satellites. The imagery spans from 1 February 2013, to 10 June 2023, covering diverse time points over an extended period. A total of 643 raw scenes were used, comprising 190 images from KOMPSAT-3 and 453 from KOMPSAT-3A.

The dataset was constructed with consideration of multiple factors to ensure usability under diverse conditions. First, regional distribution was considered by capturing urban areas, ports, airports, and other locations worldwide, with the overall visual distribution shown in Figure 1. Table 3 provides a precise quantitative breakdown of the 4003 image patches by continent. As the table indicates, data was collected from all continents except

Antarctica, though the proportion of images from Asia and Europe is relatively high, reflecting the primary acquisition strategies and areas of interest of the KOMPSAT missions. Additionally, images from various observation angles were included. Given the nature of satellite imagery, which can be influenced by external factors such as weather conditions, illumination, and atmospheric states at the time of acquisition, images were selected to reflect a variety of environmental conditions.

Table 3. Geographical distribution of the 4003 image patches by continent.

Continent	KOMPSAT-3	KOMPSAT-3A	Total Patches
Asia	237	1360	1597
Africa	48	222	270
North America	57	367	424
South America	136	204	340
Europe	171	689	860
Australia	87	431	518
Total	**733**	**3270**	**4003**

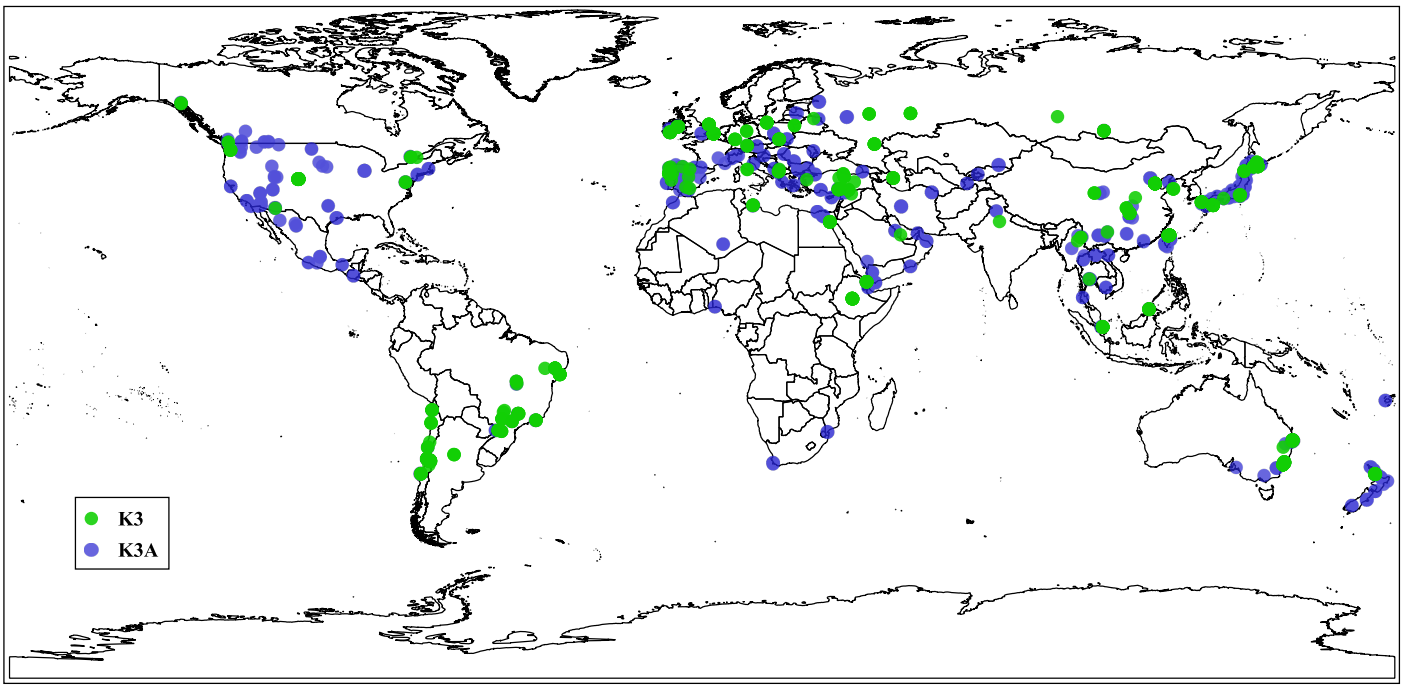


Figure 1. Global distribution of image acquisition locations from KOMPSAT-3 (K3, green) and KOMPSAT-3A (K3A, blue).

3.1.2. Image Preprocessing

Raw satellite data, due to its inherent format, size, and resolution constraints, is challenging to apply directly to deep learning models. Therefore, in this study, preprocessing was conducted to enhance image quality and normalize the data. Specifically, using Catalyst Professional software (Version 2.0; formerly PCI Geomatica) , pansharpening was performed to fuse high-resolution panchromatic images with multispectral (MS) images encompassing red (R), green (G), blue (B), and near-infrared (NIR) bands. Through this technique, the resolution of MS images from KOMPSAT-3 was significantly improved from 4 m to 0.7 m, and from 3.2 m to 0.55 m for KOMPSAT-3A, thereby securing high-quality imagery.

Pan-sharpened MS images were prepared in two formats: 8-bit PNG images for visual interpretation and analysis, and 16-bit GeoTIFF images for utilization in GIS. The 8-bit PNG images emphasized visual contrast by applying histogram stretching using only the RGB bands, while the 16-bit GeoTIFF images preserved the original digital number values to enable quantitative analysis.

Subsequently, images in both formats were segmented into 1024 × 1024 pixel patches to facilitate their use as input data for deep learning models. Each patch underwent precise geometric correction using AutoGCP (function within Catalyst Professional 2.0) , ensuring that all object instances possess accurate geographic coordinates. Through this series of preprocessing steps, a total of 4003 image patches were ultimately constructed for training and validation purposes.

3.2. Annotation Strategy

3.2.1. Annotation Protocol

All objects are initially annotated in OBB format. An OBB consists of center coordinates (x, y), width, height, and rotation angle, effectively representing the geometry of directionally oriented objects. Subsequently, to ensure compatibility with various object detection frameworks and enhance user convenience, the OBB information is automatically converted into four annotation formats: YOLO, COCO, DOTA, and GeoJSON.

The YOLO format stores objects in YOLO-OBB text files consisting of class ID and four vertex coordinates $(x_1, y_1, \ldots, x_4, y_4)$, enabling utilization in the continually evolving YOLO object detection framework [37]. The COCO format is stored as a JSON structure containing object information, where each object includes a unique ID, category ID, bbox field, and image ID. The bbox field incorporates HBB information in the order of minimum x, minimum y, width, and height. In this dataset, HBBs were derived by computing the minimum and maximum values along the x and y axes from OBB coordinates and incorporating them into the COCO format's bbox field, thereby ensuring compatibility with COCO-based object detection frameworks [38].

Additionally, the DOTA format converts OBBs into four vertex coordinates and provides them in text format (.txt) alongside class information [9]. The GeoJSON format structures the same OBB information as polygons in the WGS84 geographic coordinate system, facilitating GIS-based spatial analysis.

To accommodate compatibility with COCO and other HBB-based formats, HBBs were concurrently derived by calculating the minimum and maximum coordinate values along the x and y axes from OBB coordinates. Consequently, both OBB and HBB label formats are provided for each object, allowing flexible application across diverse detection models and environments.

3.2.2. Quality Control

To ensure dataset quality, a multi-stage verification process was implemented. The process comprised three stages, with independent quality assessments conducted at each stage. In the first stage, all annotations underwent a comprehensive manual inspection to identify and correct potential errors, such as annotation omissions, class misassignments, duplicate annotations, and boundary inaccuracies. This step established the fundamental accuracy and consistency of the dataset.

In the second stage, comprehensive cross-verification among annotators was performed. All annotation outcomes underwent mutual review by the annotators, encompassing thorough examination of object boundary precision, class consistency, and potential omissions. Emphasis was placed on minimizing subjective errors and human-induced inaccuracies during this phase.

Finally, the third stage involved sample-based inspections by remote sensing experts. Samples were drawn from diverse scenes, including high-density port areas and complex urban environments, with meticulous evaluation of OBB accuracy, class appropriateness, and geographic coordinate alignment. Necessary revisions were incorporated to refine the final quality.

Through this three-stage independent verification process, the dataset achieved high standards in accuracy, consistency, and usability, enabling reliable application in various detection models and practical domains.

3.3. Class Definition

This dataset comprises a total of 33 object classes, as illustrated in Figure 2, organized in a dual hierarchical structure consisting of mid-level categories and fine-grained classes.

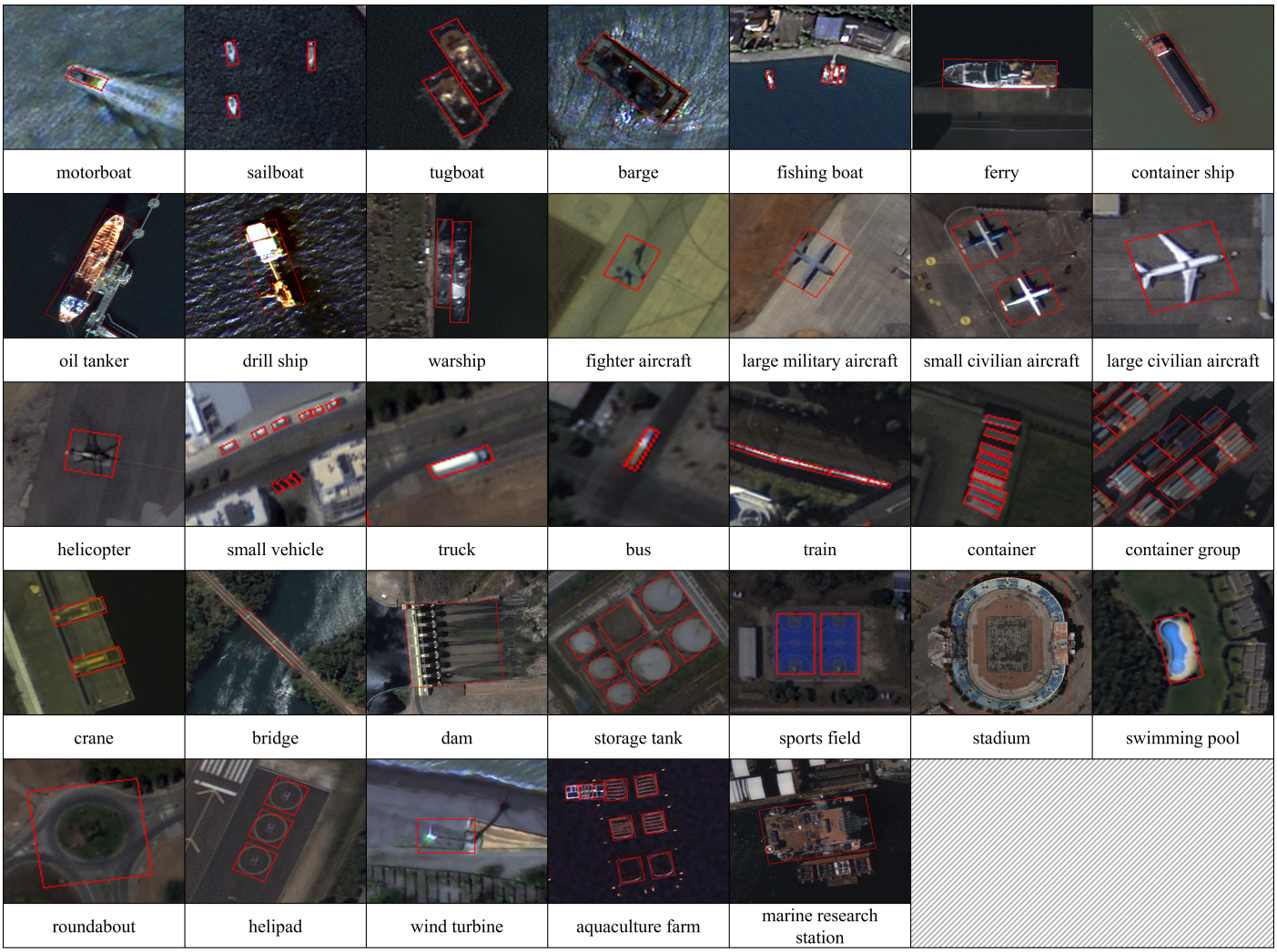


Figure 2. Example visualization of annotated object instances in diverse scenes from the dataset, demonstrating variations in object density and complexity.

The mid-level categories encompass five groups: Ship, Aircraft, Vehicle, Container, and Infrastructure. More granular object types are defined within each category (Table 4).

The Ship category includes ten fine-grained classes: motorboat (MB), sailboat (SB), tugboat (TB), barge (BG), fishing boat (FB), ferry (FR), container ship (CS), oil tanker (OT), drill ship (DS), and warship (WS).

The Aircraft category defines five classes: fighter aircraft (FA), large military aircraft (LM), small civilian aircraft (SC), large civilian aircraft (LC), and helicopter (HC).

The Vehicle category consists of four classes: small vehicle (SV), truck (TR), bus (BS), and train (TN).

The Container category is divided into container (CT) and container group (CG), while the Infrastructure category comprises 12 detailed classes: crane (CR), bridge (BR), dam (DM), storage tank (ST), sports field (SF), stadium (SD), swimming pool (SP), roundabout (RA), helipad (HP), wind turbine (WT), aquaculture farm (AF), and marine research station (MR). Through this hierarchical approach, a total of 882,399 object instances were constructed.

Table 4. All 33 classes and abbreviations in the KFGOD dataset.

Category	Class	Abbr.	Class	Abbr.
Ship	motorboat	MB	sailboat	SB
	tugboat	TB	barge	BG
	fishing boat	FB	ferry	FR
	container ship	CS	oil tanker	OT
	drill ship	DS	warship	WS
Aircraft	fighter aircraft	FA	large military aircraft	LM
	small civilian aircraft	SC	large civilian aircraft	LC
	helicopter	HC		
Vehicle	small vehicle	SV	truck	TR
	bus	BS	train	TN
Container	container	CT	container group	CG
Infrastructure	crane	CR	bridge	BR
	dam	DM	storage tank	ST
	sports field	SF	stadium	SD
	swimming pool	SP	roundabout	RA
	helipad	HP	wind turbine	WT
	aquaculture farm	AF	marine research station	MR

4. Dataset Characteristics

This section quantitatively analyzes the composition of the KFGOD dataset from various statistical perspectives. The performance of object detection models is closely tied to the distributional characteristics of the training data; thus, aspects such as class composition, object sizes, densities, instances per image, and data partitioning strategies are critical factors influencing training stability and generalization performance. Accordingly, this section systematically examines the structural features and compositional biases of the KFGOD dataset through detailed analyses of class-wise instance counts, class-wise image counts, object size, and density distributions.

4.1. Dataset Split

The dataset was split into training, validation, and test sets at an approximate 8:1:1 ratio. To ensure a reliable split that mitigates class imbalance, we implemented a stratified partitioning process.

First, the class composition and instance counts per class in each image were analyzed, and a minimum inclusion threshold was established to ensure that all classes are represented at a certain level in each dataset. This approach prevented rare classes from being concentrated in or omitted from specific datasets.

Second, the split was performed at the image level to balance the per-class instance distribution across the sets. We also ensured that images from diverse time periods

and geographical regions were evenly distributed, which promotes the model's ability to generalize.

Finally, after iterative experiments with various splitting scenarios on a total of 4003 images, the most appropriate combination was selected in terms of class coverage and balanced class distributions across datasets. Consequently, the training, validation, and test datasets comprise 3073, 470, and 460 images, respectively, with each class designed to be included above a certain threshold in all datasets. The instance counts and detailed distributions per class are summarized in Table 5, confirming that the dataset structure facilitates effective validation of model generalization capabilities during performance evaluation.

Table 5. Dataset split statistics by Class [a].

Class	MB	SB	TB	BG	FB	FR	CS	OT	DS	WS	FA	LM
Train	31,469	5536	409	1218	4231	1678	768	195	55	320	827	325
Val	3105	967	70	198	538	195	184	17	17	66	95	17
Test	5296	594	104	239	625	209	109	40	17	23	198	16
Total	39,870	7097	583	1655	5394	2082	1061	252	89	409	1120	358
Class	**SC**	**LC**	**HC**	**SV**	**TR**	**BS**	**TN**	**CT**	**CG**	**CR**	**BR**	**DM**
Train	820	1265	605	501,394	42,776	11,133	17,332	24,005	18,362	1754	497	262
Val	80	170	81	70,055	6153	1356	3712	3781	3292	283	79	47
Test	159	258	197	69,617	6441	1200	1950	4481	3294	296	83	26
Total	1059	1693	883	641,066	55,370	13,689	22,994	32,267	24,948	2333	659	335
Class	**ST**	**SF**	**SD**	**SP**	**RA**	**HP**	**WT**	**AF**		**MR**	**Total**	
Train	5486	2049	118	7982	842	989	181	1618		11	686,512	
Val	1041	325	20	1269	146	114	22	144		2	97,641	
Test	606	370	20	1100	155	165	16	357		3	98,264	
Total	7133	2744	158	10,351	1143	1268	219	2119		16	882,399	

[a] **Abbreviations**—MB: motorboat; SB: sailboat; TB: tugboat; BG: barge; FB: fishing boat; FR: ferry; CS: container ship; OT: oil tanker; DS: drill ship; WS: warship; FA: fighter aircraft; LM: large military aircraft; SC: small civilian aircraft; LC: large civilian aircraft; HC: helicopter; SV: small vehicle; TR: truck; BS: bus; TN: train; CT: container; CG: container group; CR: crane; BR: bridge; DM: dam; ST: storage tank; SF: sports field; SD: stadium; SP: swimming pool; RA: roundabout; HP: helipad; WT: wind turbine; AF: aquaculture farm; MR: marine research station.

4.2. Overall Class Distribution

The 33 defined classes exhibit variations in instance counts depending on object types, sizes, and scene characteristics. Due to differences in satellite imaging environments and object occurrence frequencies, instances are concentrated in certain classes; however, most classes include hundreds or more instances, providing sufficient distributions for training. Among all classes, SV contains the highest number of instances with approximately 640,000 labeled, while the least represented class, MR, has only 16 instances. This distribution primarily arises from real-world occurrence frequencies in high-resolution Earth observation scenes, where road- and urban-dominated tiles naturally contain many small vehicles, whereas facilities such as marine research stations are genuinely rare and geographically sparse. In addition, the very small sample size for MR also reflects practical acquisition constraints such as the sporadic availability of suitable scenes which limit the number of usable images for such rare categories. This distribution reflects real-world object occurrence patterns, and thanks to images collected from diverse regions and scenes, even rare classes secure a minimal level of diversity.

Examining by category, within the Ship category, MB has the highest count with about 40,000 instances. In the Aircraft category, LC, FA, and SC hold significant proportions. The Vehicle category's TR, BS, and TN all include over 10,000 instances, while in the Infrastructure category, SP and ST form the major classes with over 10,000 and

7000 instances, respectively. The dataset was constructed considering various time periods, regions, and scene types, ensuring that all 33 classes are substantially incorporated into the training process.

4.3. Class Frequency by Image

The distribution of object class appearances across images significantly impacts the training stability and generalization performance of detection models. In this study, based on 3073 training set images, classes were categorized into three groups—Rare, Common, and Frequent—according to the number of images in which each class appears. This stratification draws from the long-tail distribution handling strategy proposed in the Large Vocabulary Instance Segmentation (LVIS) dataset [39], serving as foundational data for systematically analyzing class imbalances and establishing effective training strategies (Table 6).

Table 6. Class-wise instance and image counts in the training set, grouped by frequency (Frequent–Common–Rare) and sorted by # Images descending.

Group	Abbr. (Class Name)	# Instances	# Images
Rare	MR (marine research station)	11	9
Common	DS (drill ship)	55	30
	AF (aquaculture farm)	1618	41
	LM (large military aircraft)	325	50
	WS (warship)	320	59
	FA (fighter aircraft)	827	89
	HC (helicopter)	605	92
Frequent	SD (stadium)	118	101
	SC (small civilian aircraft)	820	102
	WT (wind turbine)	181	103
	OT (oil tanker)	195	108
	SB (sailboat)	5536	110
	TB (tugboat)	409	140
	FB (fishing boat)	4231	158
	BG (barge)	1218	200
	CS (container ship)	768	228
	LC (large civilian aircraft)	1265	235
	DM (dam)	262	244
	TN (train)	17,332	246
	FR (ferry)	1678	250
	BR (bridge)	497	297
	CR (crane)	1754	300
	HP (helipad)	989	354
	ST (storage tank)	5486	373
	SF (sports field)	2049	418
	RA (roundabout)	842	542
	MB (motorboat)	31,469	565
	SP (swimming pool)	7982	567
	CG (container group)	18,362	571
	CT (container)	24,005	850
	BS (bus)	11,133	1127
	TR (truck)	42,776	2228
	SV (small vehicle)	501,394	2739

Rare classes are defined as those appearing in fewer than 30 images, with MR falling into this category in the dataset. With a total of 11 instances across 9 images, representing only 0.29% of all images, this sparsity poses challenges for adequate feature representation using standard training methods alone.

Common classes appear in 30 to 100 images, encompassing a total of six classes: DS, AF, LM, WS, FA, and HC. These classes exhibit moderate image appearance frequencies, with instance counts ranging from tens to thousands, enabling relatively stable training. However, limitations in scene diversity may necessitate additional regularization or balancing strategies.

Frequent classes are defined as those appearing in more than 100 images, including 26 classes such as SD, SC, WT, and OT. Notably, classes like MB, CT, BS, TR, and SV contain large numbers of instances across numerous images. In particular, SV holds 501,394 instances observed in 2739 images, accounting for approximately 89% of all images. While these classes provide abundant training data, they also carry the risk of model bias toward specific classes.

These class-wise image counts quantitatively analyze the dataset's imbalanced structure and provide crucial evidence for applying differential training strategies according to class distribution characteristics. This approach can contribute to achieving balanced recognition performance across diverse classes and minimizing training distortions due to data imbalances.

4.4. Instance Size Distribution

Object instance sizes are a critical factor influencing detection model performance, particularly in satellite imagery where objects of varying scales coexist due to resolution, scene characteristics, and object densities. In this study, all classes were categorized into four size groups—Small, Medium, Large, and Very Large—based on the average area (pixel^2) of object instances (Figure 3).

The class categorization criteria based on object sizes are commonly used in the COCO dataset, originally defined for general RGB images and HBB annotations [38]. However, satellite imagery presents fundamental limitations in directly applying COCO criteria due to high-resolution characteristics, complex background structures, varying object densities, and OBB-based annotations. Moreover, prior studies on satellite and aerial imagery object detection datasets rarely quantify or explicitly categorize object size distributions, and standardized definitions remain unestablished.

Furthermore, the dataset's actual object size statistics exhibit an extremely long-tailed distribution, with approximately 96.5% of all 880,000 instances being smaller than 500 pixel^2. A standard statistical (e.g., percentile-based) split, as used in COCO, would thus be uninformative, failing to capture the meaningful diversity of object scales. Therefore, the boundary values of 500, 5000, and 20,000 pixel^2 were selected not based on instance-level statistics but as qualitative, order-of-magnitude cutoffs to delineate semantic shifts in object type. This categorization reflects the dataset's intrinsic properties and serves as a foundational analysis for training detection models robust to diverse object scales.

The Small group consists of classes with average areas below 500 pixel^2, encompassing the largest number of instances overall. For example, the SV class holds over 640,000 instances with an average area of approximately 38 pixel^2. This group includes transportation and maritime objects such as TR, BS, CT, and MB, where the accurate detection of small objects in complex, dense clutter is critical to overall precision.

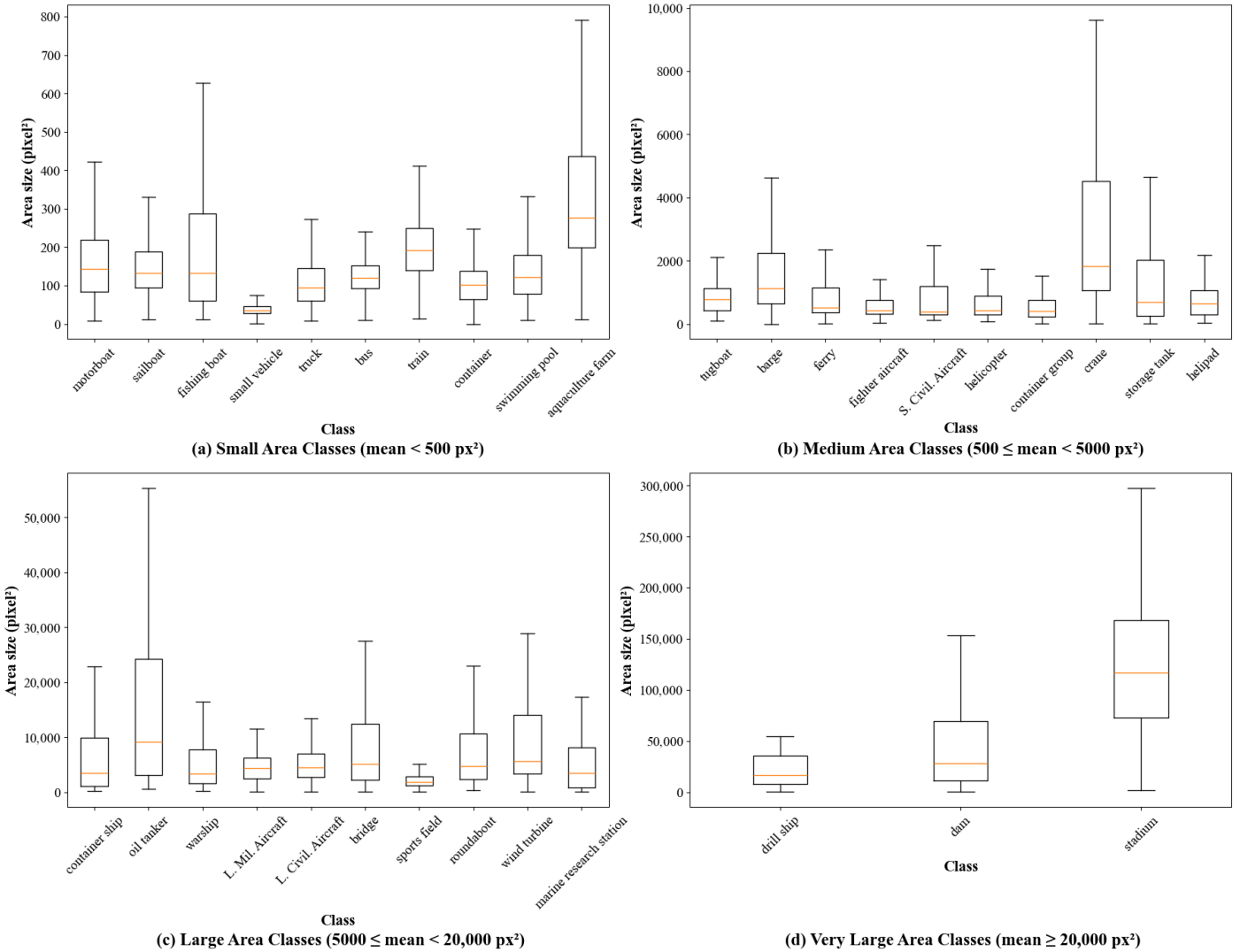


Figure 3. Instance size distribution across object classes in the dataset. The object classes are grouped into four categories based on mean area size: (**a**) small area classes, (**b**) medium area classes, (**c**) large area classes, and (**d**) very large area classes.

The Medium group features average areas between 500 and 5000 pixel^2, including diverse classes representing distinct, individual objects like FA, HC, HP, SP, and TN. Comprising medium-sized objects frequently encountered in real environments, this group enables a balanced evaluation of object identification and localization performance.

The Large group has average areas between 5000 and 20,000 pixel^2, encompassing large vessels and infrastructure representing large-scale structures such as CS, OT, WT, ST, BR, and SF. These objects are visually distinct in images but require detailed boundary recognition and background differentiation.

The Very Large group includes classes with average areas exceeding 20,000 pixel^2, such as SD, DM, and DS. The SD class, with an average area of approximately 119,000 pixel^2, can occupy an entire image patch. Such objects representing landmark-scale objects necessitate detection and boundary estimation based on global contextual information.

The composition and average areas of classes within each size group are visualized in Figure 3, intuitively illustrating relative size differences and detection difficulties among classes. This descriptive, qualitative stratification serves as practical foundational data for model design and training strategy formulation to achieve balanced detection performance across various object scales.

4.5. Instance Density per Image

The dataset exhibits diverse distributions in terms of object density, reflecting the complexity and environmental variety of real satellite imagery. The number of objects per image ranges from 0 to as many as 1942, indicating varying detection difficulties across different tiles.

Among the total 4003 images, 1645 images (approximately 41.1%) contain 100 or fewer object instances. This segment includes numerous relatively simple scenes, suitable for assessing detection performance in environments where objects are sparsely distributed over wide areas. Conversely, numerous images containing 500 or more objects provide opportunities to evaluate model generalization capabilities in congested, high-density complex scenes.

The image with the highest object count contains a total of 1942 instances. The top 20 high-density images all contain over 1300 objects, predominantly corresponding to ports, urban centers, or dense intersection areas. In contrast, 2 images contain no objects, while 193 images feature only a single object.

The overall distribution of objects per image is visualized in Figure 4, intuitively depicting scene compositions with varying densities and their frequency characteristics. This distribution is designed to enable balanced training and evaluation of model detection capabilities across both low-density and high-density scenarios.

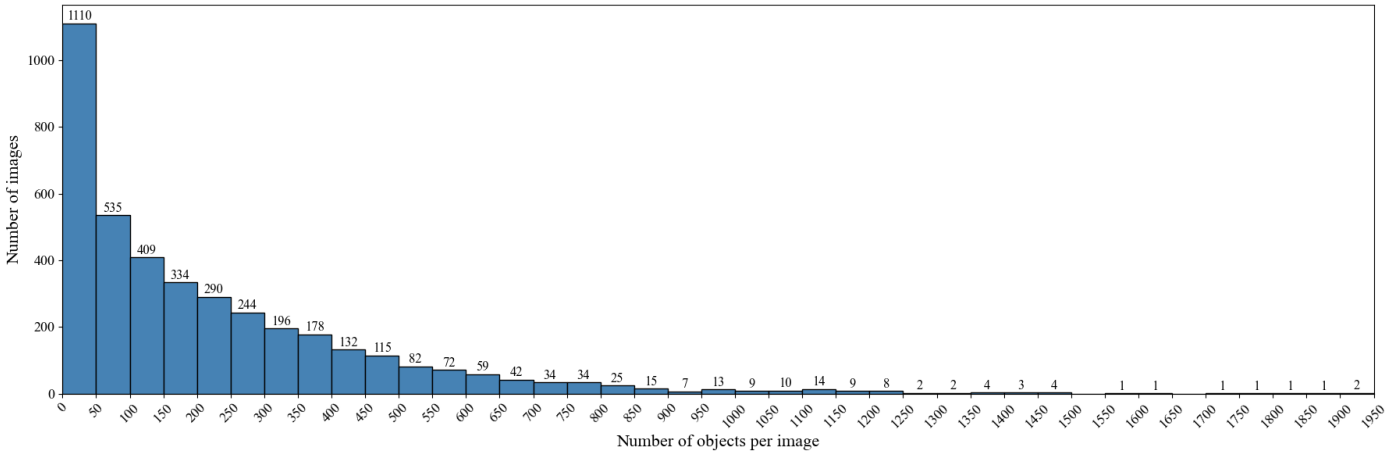


Figure 4. Distribution of the number of object instances per image in the KFGOD dataset. The histogram shows the frequency of images (y-axis) based on the number of objects they contain (x-axis), illustrating the dataset's diverse scene densities.

5. Experimental Evaluation

This section describes the various experiments conducted to validate the effectiveness of the proposed KFGOD dataset and provide baselines for future related research, along with their analysis results. To this end, state-of-the-art (SOTA) models representative of the current oriented object detection field were selected for comprehensive performance evaluation.

5.1. Experimental Setup

5.1.1. Dataset and Preprocessing

All experiments were conducted using the train, validation, and test splits of the proposed KFGOD dataset. Satellite images typically exceed thousands of pixels in dimensions, posing challenges for direct input into deep learning models due to memory constraints and training efficiency. Therefore, following the preprocessing approach in major satellite imagery object detection studies such as DOTA and FAIR1M, original images were

segmented into 1024 × 1024 pixel patches with a 256-pixel overlap (stride) to minimize object truncation.

5.1.2. Baseline Models

To objectively evaluate performance on the KFGOD dataset, various widely recognized SOTA models for oriented object detection were chosen as baselines. These models represent recent research trends, including 2-stage and 1-stage approaches, as well as CNN- and Transformer-based architectures, configured as follows.

RoI Transformer [25] and Oriented R-CNN [40]: Representative 2-stage frameworks for oriented object detection. To evaluate their performance across different backbone architectures, the publicly available mmrotate library was utilized [41]. The applied backbones include the CNN-based ResNet-101, Transformer-based Swin Transformer-Small (Swin-S), and the latest CNN architectures LSKNet and HiViT, analyzing performance variations with four backbone networks.

YOLOv11 [42]: The latest version of the YOLO series, a prominent 1-stage object detection model renowned for its speed and high performance. In this experiment, to closely analyze performance changes according to model sizes (nano, small, medium, large, and x-large), all five models of YOLOv11 officially supporting OBB were evaluated.

The various model configurations aim to demonstrate that the proposed dataset serves as a robust benchmark capable of fairly evaluating performance across diverse methodologies without bias toward specific architectures.

To ensure the reproducibility of all experiments, training parameters followed the optimal default settings recommended in each model's official implementation. For fair performance comparison, all baseline models were trained for 36 epochs uniformly.

5.1.3. Implementation Details

Key hyperparameters used for training each baseline model are detailed in Table 7. Data augmentation strategies were applied based on model type: 2-stage models (RoI Transformer and Oriented R-CNN) utilized *RResize* and *RRandomFlip*, with *PolyRandomRotate* additionally applied to LSKNet-based models. YOLOv11 models employed the default augmentation pipeline from the official ultralytics library [37], which includes *FlipLR*, *Mosaic*, and *RandAugment*.

Table 7. Key hyperparameters for baseline models.

Model	Backbone	Optimizer	Learning Rate	Total Batch Size
RoI Transformer	ResNet-101	SGD	0.005	16
	Swin-S	AdamW	0.0001	16
	HiViT	AdamW	0.0001	8
	LSKNet	AdamW	0.0001	8
Oriented R-CNN	ResNet-101	SGD	0.005	16
	Swin-S	AdamW	0.0001	16
	HiViT	AdamW	0.0001	8
	LSKNet	AdamW	0.0001	8
YOLOv11	—	AdamW	0.00027	16

As shown in Table 7, optimizers and learning rates vary between models. This was an intentional choice to ensure a fair evaluation of each architecture's optimized performance, rather than forcing arbitrary unified settings. The settings follow the official default configurations recommended in their respective repositories. For 2-stage models, the listed learning rate is the base learning rate (post-warmup), which is scheduled according to the *lr_config*.

For YOLOv11, the optimizer (AdamW) and learning rate (0.00027) were automatically calculated by the library's 'auto' setting based on the KFGOD dataset characteristics.

The only unified parameter was the training duration. For a fair comparison, all baseline models were trained for an identical 36 epochs. For the 2-stage models, this corresponds to the standard 3x schedule (36 epochs) in the mmrotate benchmark [41]. The 1-stage YOLOv11 models were also set to 36 epochs to strictly and fairly match the training duration of the 2-stage models.

For full reproducibility, the configurations used in this benchmark follow the *3x_schedule* settings in the official mmrotate repository for 2-stage models. For YOLOv11 models, the default settings from the ultralytics repository were used, with the training epochs manually set to 36.

5.2. Evaluation Tasks and Metrics

The evaluation task in this study involves accurately detecting the positions of objects in images using OBBs and classifying them into one of the 33 predefined fine-grained classes. An OBB is represented as a quadrilateral with arbitrary orientation defined by four vertex coordinates $(x_i, y_i) \mid i = 1, 2, 3, 4$.

Key metrics for quantitatively measuring detection performance adopt the Average Precision (AP) and mean Average Precision (mAP) across all classes, widely used in object detection benchmarks. AP represents the Average Precision values as recall varies from 0 to 1, calculated as the area under the precision–recall curve. Ultimately, mAP is the average of AP values over all classes, serving as an indicator of the model's overall performance on the dataset.

To determine true positives (TPs) and false positives (FPs) for the AP calculation, the Intersection over Union (IoU) metric is used to measure the overlap between a predicted bounding box (B_p) and a ground-truth bounding box (B_{gt}). In OBB detection tasks, this requires computing the intersection area of two rotated rectangles. Since B_p, B_{gt}, and their intersection ($B_p \cap B_{gt}$) are convex polygons, their areas can be calculated with standard geometric algorithms. The IoU is then defined as

$$\text{IoU} = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|} = \frac{|B_p \cap B_{gt}|}{|B_p| + |B_{gt}| - |B_p \cap B_{gt}|}$$

Following the standard evaluation protocol for oriented object detection, a prediction is classified as a TP if the IoU is 0.5 or greater; otherwise, it is considered an FP. The mAP is then computed based on this threshold.

5.3. Benchmark Results and Analysis

This section quantitatively compares model performances and conducts in-depth analysis of key characteristics based on results from applying the selected SOTA oriented object detection models to the KFGOD test dataset. Table 8 summarizes overall mAP performance for each model and backbone combination, while Table 9 details AP at the class level.

Experimental results demonstrate that the latest 1-stage detector, YOLOv11, generally outperformed traditional 2-stage models. Notably, the largest model, YOLOv11 x-large, achieved a mAP of 63.9%, surpassing all compared models, including 2-stage ones. The YOLOv11 large, medium, and small models also recorded high mAPs of 62.2%, 60.2%, and 55.4%, respectively, affirming their robust performance.

In contrast, traditional 2-stage models such as RoI Transformer and Oriented R-CNN, even with modern Transformer-based backbones (HiViT and Swin-S), plateaued at a maximum mAP of 54.4%. These findings suggest that, for the complex and diverse object

characteristics in satellite imagery, the efficient feature fusion and representation learning capabilities of the latest YOLOv11 architecture may be more effective than the RoI refinement advantages of 2-stage structures. This indicates that 1-stage architectures have reached SOTA performance in fine-grained oriented object detection, balancing accuracy and computational efficiency.

Table 8. mAP comparison of detection models with different backbones on the KFGOD test set.

Model	Backbone	mAP
RoI Transformer	ResNet101	0.467
	Swin Transformer Small	0.505
	HiViT-B	0.533
	LSKNet-S	0.522
Oriented R-CNN	ResNet101	0.507
	Swin Transformer Small	0.536
	HiViT-B	**0.544**
	LSKNet-S	0.509
YOLOv11	nano	0.482
	small	0.554
	medium	0.602
	large	0.622
	x-large	**0.639**

Table 9. Benchmark results (Average Precision, AP, in %) on the KFGOD test set grouped by coarse category with class abbreviations. The best-performing model for each class and the overall mAP for each model are highlighted in bold.

Coarse Category	Abbr.	RoI Transformer				Oriented R-CNN				YOLOv11				
		R101	Swin-S	HiViT	LSKNet	R101	Swin-S	HiViT	LSKNet	Nano	Small	Medium	Large	x-Large
Ship	MB	58.9	58.8	67.7	67.4	57.9	62.6	68.6	64.5	64.2	66.2	**72.4**	71.2	**72.4**
	SB	46.4	62.3	55.3	55.2	52.2	58.7	**67.4**	56.1	33.8	24.8	42.2	43.6	53.1
	TB	51.6	46.2	49.5	43.0	51.1	44.5	**54.6**	38.5	8.6	25.5	52.9	43.8	43.3
	BG	47.5	53.5	62.3	60.9	58.1	56.6	60.0	56.0	36.2	55.7	55.3	62.6	**63.3**
	FB	43.8	54.6	59.9	56.3	49.4	60.5	62.2	53.6	46.9	56.4	58.2	**67.0**	64.2
	FR	17.9	29.0	36.1	36.6	30.7	30.9	37.1	**39.8**	25.2	31.3	21.2	25.6	27.8
	CS	34.4	45.6	44.0	54.5	45.9	51.0	46.7	54.4	47.7	57.8	60.2	60.3	**64.8**
	OT	44.3	51.4	54.0	**66.9**	58.1	60.0	61.1	66.6	24.5	26.5	41.7	34.7	45.5
	DS	40.9	39.3	40.9	50.0	33.4	**62.4**	39.9	46.2	39.1	60.8	61.7	55.2	61.0
	WS	45.3	56.5	**74.1**	57.7	48.7	66.2	63.5	55.7	24.2	48.7	54.4	63.7	72.4
Aircraft	FA	65.7	65.5	75.2	65.0	70.8	63.5	74.9	57.0	51.8	71.5	73.1	76.1	**77.0**
	LM	23.7	18.4	20.6	23.5	20.7	25.2	18.6	11.5	28.7	36.0	30.7	**43.1**	35.5
	SC	55.1	50.4	55.9	55.0	61.0	58.1	56.7	54.8	63.5	65.8	61.8	**68.4**	67.9
	LC	84.6	81.8	82.7	85.1	85.6	84.0	81.6	87.0	**92.3**	92.0	89.8	90.9	85.4
	HC	47.6	48.9	55.1	48.8	55.6	61.0	61.9	50.3	72.7	79.5	84.3	89.7	**92.0**
Vehicle	SV	16.0	22.6	22.0	22.2	15.5	20.4	14.3	21.7	53.7	60.9	66.3	65.9	**69.0**
	TR	36.2	40.1	39.6	42.4	38.7	39.7	39.7	41.1	24.3	35.2	41.0	44.1	**47.0**
	BS	47.5	52.2	61.4	60.2	50.7	58.6	64.4	61.1	40.0	49.3	61.1	69.1	**69.6**
	TN	37.0	44.6	45.9	44.9	44.6	46.7	46.2	46.2	52.4	69.6	73.9	73.7	**75.4**
Container	CT	25.2	26.9	31.7	28.2	27.3	27.6	33.9	27.5	18.5	25.4	31.1	35.0	**35.3**
	CG	40.4	42.3	44.2	38.7	43.8	42.9	42.8	38.5	48.1	54.3	51.5	**58.2**	57.5

Table 9. *Cont.*

Coarse Category	Abbr.	RoI Transformer				Oriented R-CNN				YOLOv11				
		R101	Swin-S	HiViT	LSKNet	R101	Swin-S	HiViT	LSKNet	Nano	Small	Medium	Large	x-Large
Infrastructure	CR	23.2	36.5	34.9	31.8	23.0	29.9	32.7	30.6	21.9	34.0	41.7	41.6	**51.7**
	BR	39.3	42.9	40.1	34.0	49.7	42.4	46.8	40.5	60.3	61.1	65.0	**66.1**	66.0
	DM	30.2	33.6	37.3	30.4	39.9	32.6	43.2	33.2	43.8	45.3	43.7	**46.8**	46.7
	ST	71.2	71.4	71.6	71.4	71.1	70.7	71.4	70.8	79.0	83.8	**87.8**	87.6	87.4
	SF	57.8	57.5	54.8	54.3	60.6	61.7	55.4	53.1	55.2	65.6	71.3	**74.6**	73.5
	SD	69.0	70.1	73.2	**84.5**	81.2	79.7	79.9	82.5	78.0	72.9	81.7	76.9	80.4
	SP	63.2	63.5	63.7	68.7	62.1	64.4	64.8	63.4	64.4	73.5	77.5	78.8	**81.6**
	RA	90.3	90.2	89.9	89.0	90.1	87.8	89.7	90.3	92.2	97.0	96.5	**97.9**	97.2
	HP	60.6	57.7	62.8	62.1	61.4	65.8	63.9	55.5	51.0	59.3	68.0	72.7	**77.4**
	WT	54.5	71.7	72.7	54.5	54.5	70.7	70.2	49.4	65.4	55.8	70.5	**78.1**	76.3
	AF	72.0	79.8	81.4	80.5	80.9	80.4	81.3	81.1	82.7	86.6	**89.8**	87.1	89.5
	MR	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	**7.7**	3.6	1.2
mAP		46.7	50.5	53.3	52.2	50.7	53.6	54.4	50.9	48.2	55.4	60.2	62.2	**63.9**

As evident in Table 9, visually distinct classes such as RA (up to 97.9%) and LC (up to 92.3%) achieved high APs, whereas substantial performance drops were observed in numerous classes. To deeply investigate the causes of these variations, a confusion matrix (Figure 5) was analyzed, focusing on prediction results from the top-performing YOLOv11 x-large model, to identify primary error patterns.

Confusion matrix analysis revealed that the primary cause of performance degradation is misclassification among visually similar classes. In the Ship category, confusions among similar-shaped classes were prominent, with asymmetric patterns between SB and MB: of 673 actual SB instances, 204 (30.3%) were mispredicted as MB, whereas only 209 (8.5%) of 2451 MB instances were misclassified as SB, indicating greater difficulty in learning the SB unique features.

This trend appeared more complex in the Vehicle category. Among 3188 actual TR instances, 266 (8.3%) were misclassified as SV and 128 (4.0%) as BS. Additionally, of 1310 BS instances, 91 (6.9%) were predicted as TR and 81 (6.2%) as CT, demonstrating misclassifications among multiple classes with similar rectangular forms, highlighting the limitations in distinguishing fine subtypes.

Beyond subtle shape differences between classes, physical sizes or functional similarities also emerged as key confusion factors. In the Aircraft category, distinguishing military from civilian aircraft proved challenging, with LM showing one of the highest misclassification rates: 10 (29.4%) of 34 instances mispredicted as LC. This directly explains the unusually low AP of 35.5% for LM in Table 9, underscoring that learning minute differences like aircraft sizes or wing shapes is a core challenge in fine-grained recognition.

Alongside inter-class misclassifications, failure to detect objects altogether treating them as background (false negatives) was another major error type, particularly prevalent in rare classes with fewer instances. For instance, 24 (37.5%) of 64 TB instances and 255 (37.9%) of 673 SB instances went undetected. In the most extreme case, the MR class was largely undetected across models, clearly highlighting the long-tail distribution issue as a key challenge to address.

These various error patterns, including asymmetric confusions between subclasses, misclassifications based on functional/morphological similarities, and high detection failure rates for rare classes affirm that KFGOD is a high-difficulty benchmark demanding advanced capabilities beyond mere object presence detection, including precise fine-grained visual recognition. This provides concrete directions for future algorithm im-

provements, with Figure 6 visually illustrating these key analyses alongside representative detection examples.

Figure 7 visually corroborates the quantitative findings from the confusion matrix, providing explicit failure cases. The example in Figure 7a highlights the challenge of high inter-class similarity in dense maritime scenes; the ambiguous, small rectangular profiles of MB and SB instances lead to frequent misclassifications (orange boxes), supporting the asymmetric confusion pattern observed in the matrix. Similarly, Figure 7b demonstrates the dual challenges of inter-class similarity and data sparsity. The misclassification of an LM as an LC (orange box) underscores the difficulty in distinguishing aircraft based on subtle morphological differences, such as wing shape or fuselage size. Concurrently, the missed detections (red boxes) confirm the model's struggle with the rare LM class, reinforcing the long-tail challenge.

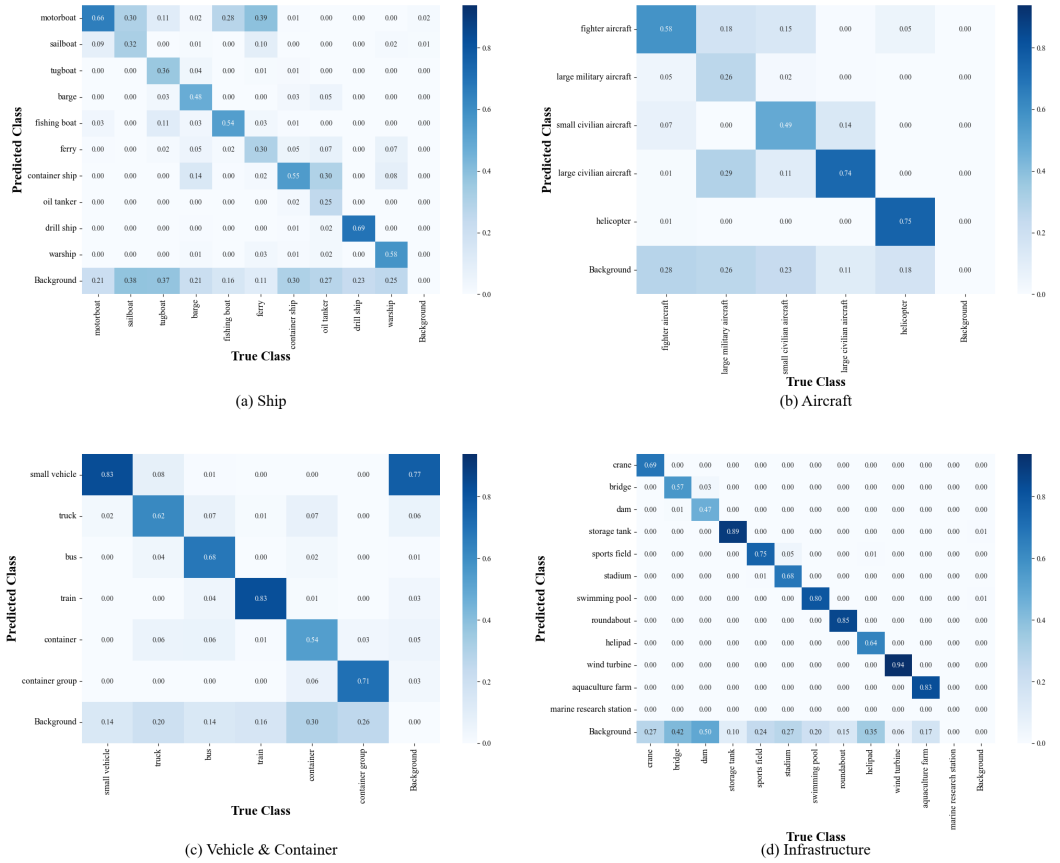
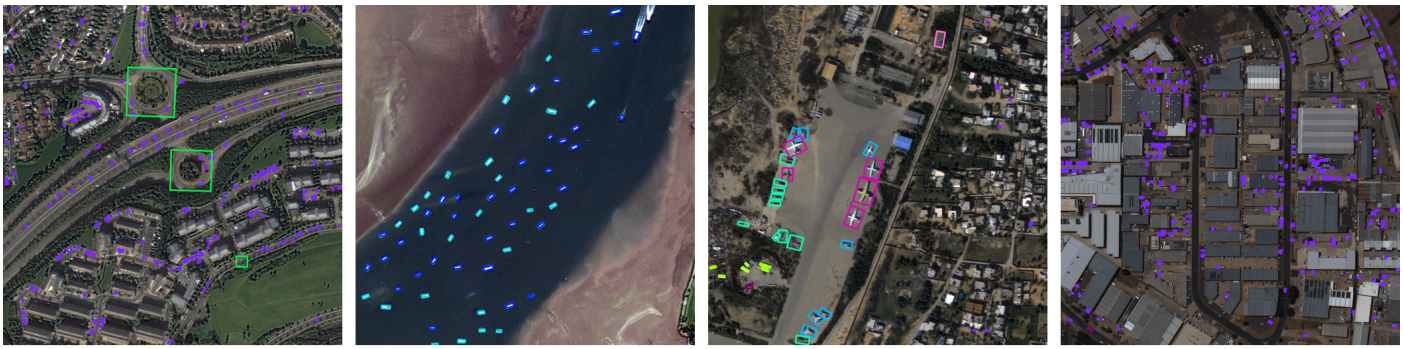


Figure 5. Normalized confusion matrices for the YOLOv11x-obb model, categorized by class type. The figure is divided into four main categories: (**a**) Ship, (**b**) Aircraft, (**c**) Vehicle and Container, and (**d**) Infrastructure. This breakdown allows for a more detailed analysis of the model's performance and confusion patterns within each specific category.

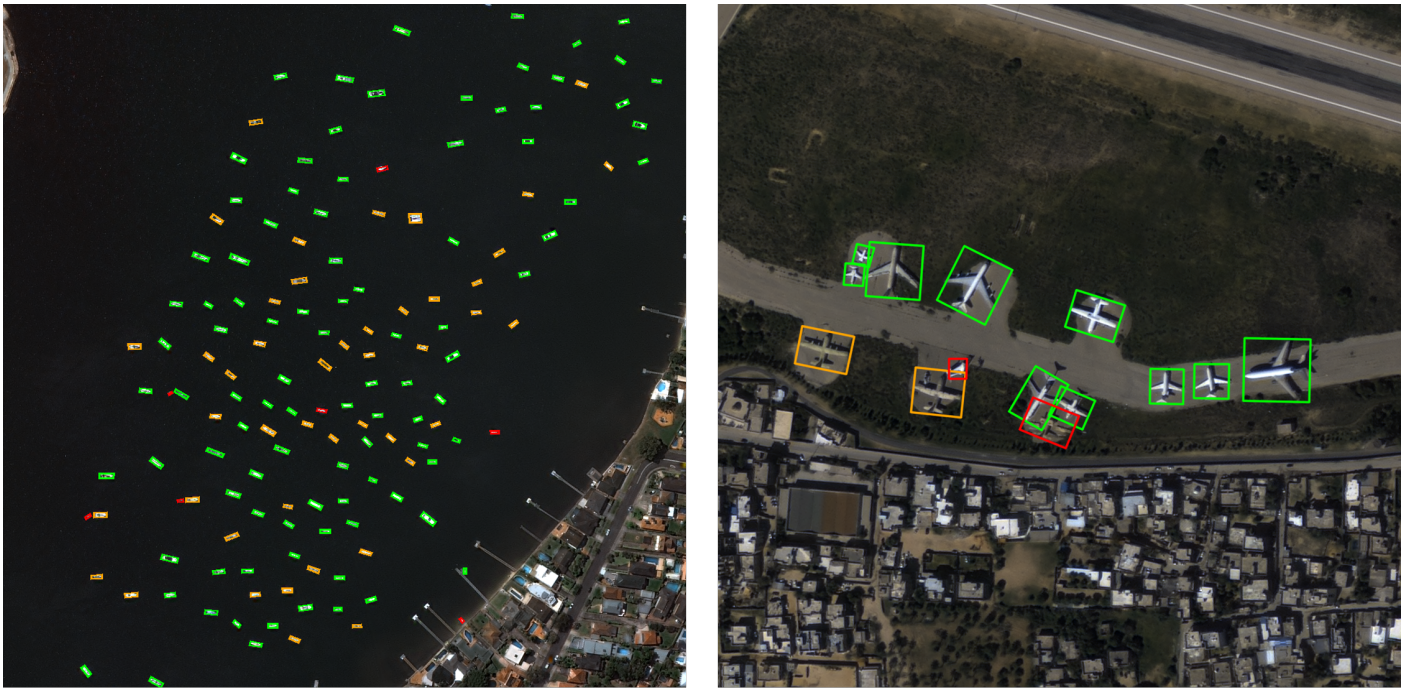
Finally, to investigate the impact of annotation format on detection performance, an additional experiment was conducted using HBB annotations with the top-performing YOLOv11 x-large model. As detailed in Table 10, the model achieved an overall mAP@0.50 of 56.2%. This result is significantly lower than the 63.9% achieved with OBBs (Table 9), representing a relative performance decrease of approximately 12%.

This performance degradation was particularly pronounced for elongated or diagonally oriented object classes. For instance, the AP for CR dropped sharply from 51.7% with the OBB model to 30.7% with the HBB model. Similarly, BR saw a significant decline from 66.0% to 37.9%, and a performance reduction was also observed for ship types like CS, which fell from 64.8% to 53.5%.



(a) (b) (c) (d)

Figure 6. Analysis of key detection cases in the KFGOD dataset. (**a**) High detection performance on the structurally clear RA class. (**b**) Potential for misclassification due to visual similarity between small ship types such as MB and SB. (**c**) Example of the detection difficulty of the uniquely shaped LM class, in contrast to LC. (**d**) Examples of detection success and failure (misses) for dense SV objects.



(a) (b)

Figure 7. Detailed visual analysis of key fine-grained confusion pairs. (**a**) A dense maritime scene illustrating confusion between MB and SB. (**b**) An airfield scene showing detection challenges for LM. Bounding box colors indicate the detection result: Green = Correct Detection, Orange = Misclassification, Red = Missed Detection.

Conversely, large objects with less directional ambiguity maintained robust performance with HBBs, showing little difference from the OBB results. For example, the performance for LC actually increased slightly from 85.4% (OBB) to 90.0% (HBB), and TN also improved from 75.4% (OBB) to 81.5% (HBB). While ST (87.4% OBB, 78.7% HBB) and AF (89.5% OBB, 82.4% HBB) experienced minor performance drops, they still recorded high AP values. The rarest class, MR, performed extremely poorly in both models (1.2% OBB, 2.1% HBB), suggesting that its difficulty stems more from data scarcity than annotation format.

In summary, while HBBs can perform adequately for certain object types, these results empirically confirm that OBB annotations are essential for tasks requiring the fine-grained discrimination of slender or densely packed objects.

Table 10. Benchmark results using horizontal bounding boxes (HBBs) on the KFGOD test set. The table shows per-class AP (%) and the overall mAP (%) achieved by the YOLOv11 x-large model.

Class	**MB**	**SB**	**TB**	**BG**	**FB**	**FR**	**CS**	**OT**	**DS**	**WS**	**FA**	**LM**
AP	65.9	49.8	44.4	63.0	56.0	41.2	53.5	47.8	31.4	52.4	74.7	43.7
Class	**SC**	**LC**	**HC**	**SV**	**TR**	**BS**	**TN**	**CT**	**CG**	**CR**	**BR**	**DM**
AP	63.6	90.0	73.3	67.4	48.6	57.1	81.5	39.5	41.8	30.7	37.9	38.1
Class	**ST**	**SF**	**SD**	**SP**	**RA**	**HP**		**WT**	**AF**		**MR**	**mAP**
AP	78.7	63.7	68.8	63.5	68.1	61.2		73.9	82.4		2.1	**56.2**

6. Discussion

This study presents the large-scale fine-grained object detection dataset KFGOD based on KOMPSAT-3/3A satellite imagery and conducts comprehensive benchmarking experiments using SOTA models. This section examines the academic value and inherent characteristics of the KFGOD dataset revealed through experimental results, highlights its practical value via real-world application cases, and finally discusses the study's limitations and future research directions.

The benchmarking experiments demonstrate that KFGOD poses a highly challenging task even for current SOTA models. The 63.9% mAP achieved by the top-performing YOLOv11 x-large model, while substantially outperforming 2-stage models, still leaves considerable room for improvement. This positions KFGOD alongside other large-scale benchmarks such as DOTA and FAIR1M as a complex, high-difficulty resource yet to be fully conquered in academia.

Through benchmarking outcomes, the dataset's core challenges stem from two properties of fine-grained classes: inter-class similarity and intra-class variance, leading to classification issues. The former manifests in misclassifications among CS and OT, BS and TR, and numerous Ship classes as confirmed in the confusion matrix analysis in Section 5.3. The latter is evident in the markedly lower AP of FA compared to similarly sized LC, attributed to the high morphological diversity from various models and imaging times included in FA. These traits demand models that can discern subtle visual cues beyond mere detection, signifying KFGOD as an ideal testbed for fine-grained recognition algorithm research.

Notably, confidently attributing performance degradations to intrinsic data properties like 'inter-class similarity' is enabled by this study's key contribution: sensor consistency. In prior heterogeneous sensor datasets, distinguishing whether performance issues arise from algorithmic limitations or domain gaps between sensors was unclear. In contrast, KFGOD fundamentally controls the sensor variable, clarifying that observed errors stem purely from algorithmic performance limits. This provides an environment for undistorted measurement of algorithm improvements, holding significant academic value.

Beyond its academic role, KFGOD has already demonstrated practical value in real-world applications by South Korean national agencies. Conventional medium-to-large satellites suffer from long revisit cycles, hindering immediate responses to maritime emergencies. To address this, the Korea Coast Guard is developing a real-time search and rescue system using microsatellite constellations capable of short-cycle imaging in specific regions, where ship detection models trained on this dataset play a pivotal role. Through response drills in actual sea areas, successfully detecting and rapidly providing location information for 2-ton small drifting vessels from urgently acquired satellite imagery contributed to securing golden time for life-saving operations [43]. This exemplifies the high generalization

performance and reliability of KFGOD, as well as its effective applicability to imagery from new satellite platforms.

Furthermore, KFGOD is being applied to the product analysis system for the Newspace Earth Observation Satellite (NEONSAT) microsatellite constellation under development by KARI. Serving as reference data for developing core functionalities to automatically detect key objects of interest from satellite imagery, it contributes to establishing foundational technologies for national satellite information utilization systems [44].

Extending beyond direct object detection, KFGOD's precise annotation information expands its value as a foundational dataset for training satellite imagery-specialized large multimodal models (LMMs). In independent research, fine-tuning an LMM using a question-answering dataset combining this dataset's detection labels and metadata achieved image understanding capabilities surpassing GPT-4o [45].

Despite its contributions, this study harbors several limitations. First, the total image count of 4003 may be quantitatively smaller compared to some recent datasets like FAIR1M or DIOR. However, this stems from a deliberate design prioritizing quality over quantity to ensure data reliability and consistency. KFGOD exclusively uses KOMPSAT-3/3A series imagery to eliminate domain gap issues from heterogeneous sensors, a strategic choice to provide a clean benchmark for fair, undistorted algorithmic performance evaluation. It is crucial to note that KFGOD is not intended to replace these valuable heterogeneous benchmarks but rather to act as a vital complement. It serves a dual role: first, as the 'controlled lab' discussed, for fair algorithmic comparison; and second, as a high-quality VHR (0.55–0.7 m) data source to address a practical resource gap. By combining KFGOD with existing heterogeneous datasets, researchers can achieve a powerful synergy: models can learn fine-grained details from the consistent data of KFGOD while simultaneously learning real-world robustness from the varied sensors and resolutions in other benchmarks.

Second, the benchmarking scope of this study is intentionally foundational. Widely used SOTA models (e.g., Faster R-CNN and YOLO series) were selected to establish a robust and reproducible baseline. Consequently, this initial study does not include evaluations of more specialized algorithms designed specifically for fine-grained object detection, such as those focused on learning highly discriminative representations [46].

Third, severe class imbalances with a long-tail distribution persist, with data concentrated in specific classes like SV. This, as evidenced by near-zero AP for MR in experiments, renders model training exceedingly difficult for rare classes with few samples. To mitigate this long-tailed distribution in practice, two lightweight remedies are recommended that require minimal code changes: (i) class-balanced re-weighting based on the effective number of samples [47] to reduce the over-dominance of frequent classes; and (ii) class-aware (repeat-factor) sampling as in LVIS [39] to upsample tail categories during training. As an optional alternative when easy negatives overwhelm dense detectors, focal loss can also stabilize learning by down-weighting well-classified examples [14].

Nevertheless, these limitations revealed in the study offer meaningful directions for future algorithmic research utilizing KFGOD. For instance, the high confusion among fine-grained classes confirmed in experiments underscores the need for sophisticated representation learning to distinguish minute visual differences between classes. This directly links to another high-priority future direction: applying and evaluating the more specialized fine-grained detection algorithms mentioned in the identified limitations. KFGOD can be utilized as a benchmark for these advanced models, particularly those designed to learn discriminative representations, to systematically assess their effectiveness against the core challenges of inter-class similarity and intra-class variance that have been identified. Additionally, low detection performance for rare classes like MR highlights the importance of few-shot or long-tail learning techniques effective with minimal data. Lastly,

detection challenges for small, dense objects such as SV can lead to specialized architecture development research.

Long-term plans are also in place to continually advance KFGOD. In the short term, augmenting data for rare classes to mitigate imbalances and acquiring time-series imagery for change detection utility are planned. In addition, targeted acquisition is planned for scenes that are likely to contain tail categories (e.g., additional coastal research complexes for *MR* and selected military airfields for *LM*) and periodic updates will be released after curation and quality checks; future releases will include a changelog documenting newly added tail-class imagery. Through these efforts, KFGOD will evolve into an even more valuable and comprehensive resource for satellite imagery analysis research.

7. Conclusions

In this paper, to expand high-resolution satellite imagery object detection research, we propose KFGOD, a large-scale fine-grained benchmark dataset based on KOMPSAT-3/3A imagery, unaddressed in existing datasets. Comprising 33 classes and approximately 880,000 OBB-annotated instances, KFGOD embodies realistic satellite imagery challenges, including fine-grained class distinctions, extreme scale variations, and density imbalances. Benchmarking with SOTA models revealed classification issues among similar classes as the primary hurdle, providing key directions for future algorithm development.

Furthermore, the deployment of KFGOD in the Korea Coast Guard's maritime surveillance system and microsatellite constellation utilization system demonstrates its value as a practical resource supporting national missions beyond academic benchmarking.

In the future, we plan to continually enhance the KFGOD dataset through sparse class data augmentation and time-series dataset construction for change detection. Thereby, we aim to contribute to technological advancements in Earth observation and anticipate this dataset serving as a meaningful resource for researchers in related fields.

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

Funding: This study was conducted with the support of the "Development of Application Support System for Satellite Information Big Data(RS-2022-00165154)" of the Korea Aerospace Administration.

Data Availability Statement: The KOMPSAT Fine-Grained Object Detection (KFGOD) dataset presented in this study is publicly available at the Korea Research Data Platform (DataON) at https://doi.org/10.22711/idr/1101.

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

References

1. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [CrossRef]
2. Gui, S.; Song, S.; Qin, R.; Tang, Y. Remote sensing object detection in the deep learning era—A review. *Remote Sens.* **2024**, *16*, 327. [CrossRef]
3. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
4. Lam, D.; Kuzma, R.; McGee, K.; Dooley, S.; Laielli, M.; Klaric, M.; Bulatov, Y.; McCord, B. xView: Objects in Context in Overhead Imagery. *arXiv* **2018**, arXiv:1802.07856. [CrossRef]
5. Christie, G.; Fendley, N.; Wilson, J.; Mukherjee, R. Functional map of the world. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6172–6180. [CrossRef]

6. Minetto, R.; Segundo, M.P.; Rotich, G.; Sarkar, S. Measuring human and economic activity from satellite imagery to support city-scale decision-making during covid-19 pandemic. *IEEE Trans. Big Data* **2020**, *7*, 56–68. [CrossRef]
7. Chen, Y.; Qin, R.; Zhang, G.; Albanwan, H. Spatial temporal analysis of traffic patterns during the COVID-19 epidemic by vehicle detection using planet remote-sensing satellite images. *Remote Sens.* **2021**, *13*, 208. [CrossRef]
8. Golej, P.; Horak, J.; Kukuliac, P.; Orlikova, L. Vehicle detection using panchromatic high-resolution satellite images as a support for urban planning. Case study of Prague's centre. *GeoScape* **2022**, *16*, 108–119. [CrossRef]
9. Ding, J.; Xue, N.; Xia, G.S.; Bai, X.; Yang, W.; Yang, M.Y.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7778–7796. [CrossRef] [PubMed]
10. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 116–130. [CrossRef]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
14. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
15. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [CrossRef]
16. Jakubik, J.; Roy, S.; Phillips, C.; Fraccaro, P.; Godwin, D.; Zadrozny, B.; Szwarcman, D.; Gomes, C.; Nyirjesy, G.; Edwards, B.; et al. Foundation models for generalist geospatial artificial intelligence. *arXiv* **2023**, arXiv:2310.18660. [CrossRef]
17. Tong, K.; Wu, Y.; Zhou, F. Recent advances in small object detection based on deep learning: A review. *Image Vis. Comput.* **2020**, *97*, 103910. [CrossRef]
18. Cheng, G.; Yuan, X.; Yao, X.; Yan, K.; Zeng, Q.; Xie, X.; Han, J. Towards large-scale small object detection: Survey and benchmarks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13467–13488. [CrossRef]
19. Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 1452–1459. [CrossRef]
20. Shi, F.; Zhang, T.; Zhang, T. Orientation-aware vehicle detection in aerial images via an anchor-free object detection approach. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5221–5233. [CrossRef]
21. Wang, J.; Yang, W.; Guo, H.; Zhang, R.; Xia, G.S. Tiny object detection in aerial images. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3791–3798. [CrossRef]
22. Zhu, P.; Wen, L.; Du, D.; Bian, X.; Fan, H.; Hu, Q.; Ling, H. Detection and tracking meet drones challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7380–7399. [CrossRef]
23. Li, X.; Men, F.; Lv, S.; Jiang, X.; Pan, M.; Ma, Q.; Yu, H. Vehicle detection in very-high-resolution remote sensing images based on an anchor-free detection model with a more precise foveal area. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 549. [CrossRef]
24. Liu, Z.; Wang, H.; Weng, L.; Yang, Y. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1074–1078. [CrossRef]
25. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI transformer for oriented object detection in aerial images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2849–2858. [CrossRef]
26. Rufener, M.C.; Ofli, F.; Fatehkia, M.; Weber, I. Estimation of internal displacement in Ukraine from satellite-based car detections. *Sci. Rep.* **2024**, *14*, 31638. [CrossRef]
27. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [CrossRef]
28. Mundhenk, T.N.; Konjevod, G.; Sakla, W.A.; Boakye, K. A large contextual dataset for classification, detection and counting of cars with deep learning. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part III 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 785–800. [CrossRef]
29. Chen, K.; Wu, M.; Liu, J.; Zhang, C. FGSD: A dataset for fine-grained ship detection in high resolution satellite images. *arXiv* **2020**, arXiv:2003.06832. [CrossRef]
30. Cheng, G.; Zhou, P.; Han, J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]

31. Azimi, S.M.; Bahmanyar, R.; Henry, C.; Kurz, F. Eagle: Large-scale vehicle detection dataset in real-world scenarios using aerial imagery. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 6920–6927. [CrossRef]
32. Al-Emadi, N.; Weber, I.; Yang, Y.; Ofli, F. VME: A Satellite Imagery Dataset and Benchmark for Detecting Vehicles in the Middle East and Beyond. *Sci. Data* **2025**, *12*, 500. [CrossRef]
33. Higuchi, A. Toward more integrated utilizations of geostationary satellite data for disaster management and risk mitigation. *Remote Sens.* **2021**, *13*, 1553. [CrossRef]
34. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 3735–3739. [CrossRef]
35. Zhang, Y.; Yuan, Y.; Feng, Y.; Lu, X. Hierarchical and Robust Convolutional Neural Network for Very High-Resolution Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5535–5548. [CrossRef]
36. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [CrossRef]
37. Jocher, G.; Qiu, J. *Ultralytics YOLO11*, version 11.0.0; Ultralytics: 2024. Available online: https://github.com/ultralytics/ultralytics (accessed on 17 November 2025)
38. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part v 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755. [CrossRef]
39. Gupta, A.; Dollar, P.; Girshick, R. LVIS: A Dataset for Large Vocabulary Instance Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [CrossRef]
40. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3520–3529. [CrossRef]
41. Zhou, Y.; Yang, X.; Zhang, G.; Wang, J.; Liu, Y.; Hou, L.; Jiang, X.; Liu, X.; Yan, J.; Lyu, C.; et al. Mmrotate: A rotated object detection benchmark using pytorch. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10–14 October 2022; pp. 7331–7334. [CrossRef]
42. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
43. Lee, D.H.; Choi, K. Small Ship Detection for Marine Search and Rescue: Construction of High-Resolution Training Data and Performance Evaluation Using Microsatellite Imagery. *Korean J. Remote Sens.* **2024**, *40*, 943–955. [CrossRef]
44. Kim, H.O.; Ha, J.S.; Jeong, S.; Kim, Y.; Park, S.; Oh, H. High Resolution Land Application Approach Using Micosatellite Costellation (NEONSAT) in Korea. In Proceedings of the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 7168–7170. [CrossRef]
45. Oh, H.; Shin, D.B.; Chung, D.W. KOMPSAT-3/3A Image-text Dataset for Training Large Multimodal Models. *GEO DATA* **2025**, *7*, 27–35. [CrossRef]
46. Xie, X.; Cheng, G.; Li, W.; Lang, C.; Zhang, P.; Yao, Y.; Han, J. Learning Discriminative Representation for Fine-Grained Object Detection in Remote Sensing Images. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 8197–8208. [CrossRef]
47. Cui, Y.; Jia, M.; Lin, T.Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9260–9269. [CrossRef]