

EarthNets: Empowering AI in Earth Observation

Zhitong Xiong, *Member, IEEE*, Fahong Zhang[†], Yi Wang[†], Yilei Shi, *Member, IEEE*, and Xiao Xiang Zhu^{*}, *Fellow, IEEE*

Abstract—Earth observation, aiming at monitoring the state of planet Earth using remote sensing data, is critical for improving our daily lives and living environment. With an increasing number of satellites in orbit, more and more datasets with diverse sensors and research domains are published to facilitate the research of the remote sensing community. In this paper, for the first time, we present a comprehensive review of **more than 400** publicly published datasets, including applications like, land use/cover, change/disaster monitoring, scene understanding, agriculture, climate change and weather forecasting. We systematically analyze these Earth observation datasets from five aspects, including the volume, bibliometric analysis, research domains and the correlation between datasets. Based on the dataset attributes, we propose to measure, rank and select datasets to build a new benchmark for model evaluation. Furthermore, a new platform for Earth observation, termed EarthNets, is released towards a fair and consistent evaluation of deep learning methods on remote sensing data. EarthNets supports standard dataset libraries and cutting-edge deep learning models to bridge the gap between remote sensing and machine learning community. Based on the EarthNets platform, extensive deep learning methods are evaluated on the new benchmark. The insightful results are beneficial to future research. The platform, dataset collections are publicly available at <https://earthnets.nicepage.io/>.

Index Terms—Disentangled representation learning, height estimation, interpretable deep models, remote sensing, Transformer deep networks

1 INTRODUCTION

EARTH Observation (EO) aims to monitor and assess the status of the Earth’s surface using various remote sensing (RS) technologies [1], [2]. EO can make a significant contribution to our ability to better understand and analyse the planet Earth using the RS data. The research in EO has been successfully applied to urban planning [3], natural resources management [4], agriculture [5], food security [6] and disaster monitoring [7], [8]. All these applications have great potential for serving humanity.

With the development of Earth observation technology, more and more satellites with diverse imaging sensors are launched for different missions. To enable a broad range of real-world applications, a huge amount of RS data with global-coverage and high-resolution is received per day for automatic processing and analysis. To deal with large-scale data, deep learning techniques [9] have been proven effective for many different research areas. One of the key success of deep learning methods is to train model with large-scale data. In this context, recent RS datasets are constructed with larger and larger volumes. In Fig. 1, we show a chronological overview of the volumes of existing 401 datasets. As seen from this figure, more and more, larger and larger datasets have been constructed and published during the past decade.

With the overwhelming success of deep learning techniques, there have been quite a few deep models developed for different tasks on RS data [10], [11]. Although considerable progress has been made and promising results on many RS datasets have been achieved, there are still several problems that need more research efforts to handle.

There is a lack of comprehensive dataset review for EO tasks. Owing to the efforts made by EO researchers, there are numerous datasets in the RS community with different modalities, resolutions, and application domains. RS data can be in different modalities, including optical (RGB), hyperspectral, synthetic aperture radar (SAR), multi-spectral (MS), point cloud, and so forth. Regarding the application domains, the published datasets may be designed for land use and land cover, change monitoring [12], disaster monitoring [13], scene recognition, semantic segmentation, ground object detection, object tracking, agriculture, climate change and weather forecasting. A comprehensive review of the RS datasets can provide researchers a holistic view of the status of the research community. Although a few of works [14], [15], [16] try to review the RS datasets, there is still no exhaustive and comprehensive review for them.

There is no systematic summary and analysis of RS datasets. The ever-growing quantity of RS datasets makes it difficult to find the proper one for a specific application. For example, there are more than 20 datasets related to building extraction [17]. It could be laborious to find and select the most suitable one for the desired application. Thus, it is vitally important to summarize and categorize RS datasets to provide valuable guidance and reference for researchers. Moreover, insightful analysis of existing datasets can help researchers understand the current research state and trends of the whole community. Systematically analyzing RS datasets can provide guidance to future research

- Z. Xiong, F. Zhang, Y. Wang and X. X. Zhu are with the Chair of Data Science in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany.
- Y. Shi is with the Chair of Remote Sensing Technology, Technical University of Munich (TUM), 80333 Munich, Germany.
- [†] indicates equal contribution, and ^{*} indicates the corresponding author (e-mail: xiaoxiang.zhu@tum.de).

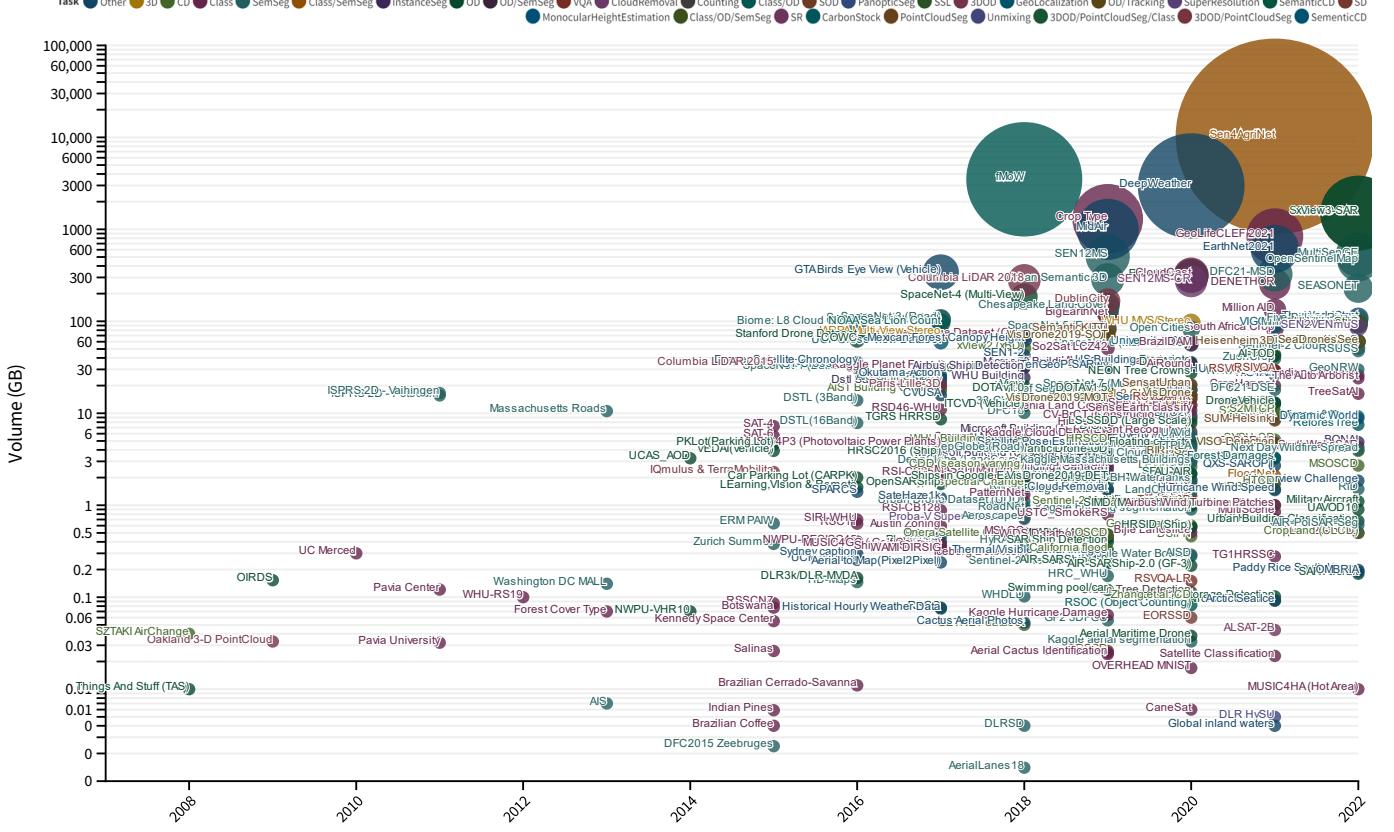


Fig. 1. Chronological overview of the volumes of existing 400 datasets. As we can see, more and more, larger and larger datasets have been constructed and published during the past decade. (Best viewed with zoom in.)

directions.

There is no unified benchmark for fairly comparing remote sensing methods. In the computer vision (CV) community, some large-scale dataset like ImageNet [18] is usually used for the evaluation of newly developed deep learning models. Compared with small-scale datasets, large-scale ones with rich semantic annotations align better with the complex real-world scenarios [14]. Thus they can be more reliable for the performance validation and comparison of deep learning algorithms. Although several RS datasets with large volumes are published [19], [20], [21], many of the currently developed methods are still evaluated on small-scale datasets. However, datasets with small-scale or limited geographic coverage may bias to a specific data distribution rather than the real-world scenarios. Moreover, many RS datasets are published with no standard train/validation/test splits. It increases the uncertainty during the evaluation of algorithms. Thus, it is urgent to build new RS benchmarks to enable a fair comparison of different methods.

There is a lack of an open platform for different EO tasks. For deep learning methods, the backbone networks, hyper-parameters and training tricks are influential factors that should be considered for fair performance comparison. However, existing works usually evaluate the performance with different dataset splits, which makes it difficult to fairly and reliably compare different algorithms. Due to the large variance in data collection sensors and pre-processing

pipelines, it is non-trivial to directly adapt modern deep learning models to RS datasets [22]. As a result, lots of cutting-edge and off-the-shelf deep learning methods from the machine learning community are not evaluated and compared on RS data. To handle the aforementioned problems, in this study, we first make an exhaustive and comprehensive review of the publicly accessible RS datasets. Specifically, we review more than 400 datasets, and provide detailed attribute information for each dataset. Next, a systematic analysis is made based on the information of these datasets. Based on the attribute information, we filter, rank and select five large-scale datasets designed for general purpose to build a new benchmark for model evaluation. This new benchmark contains datasets of different resolutions, research tasks, and data modalities. To further enable a fair and reproducible comparison of different algorithms, we construct a new deep learning platform, termed EarthNets, for future work to build on.

To sum up, our main contributions can be summarized as follows.

- 1) We review more than 400 datasets published in the RS community. These datasets are summarized and categorized into different research tasks and research domains. Detailed attributes including ten different aspects are provided for each dataset.
- 2) Systematic analyses are made from five aspects to provide insights and ideas for the future research of the RS community. Specifically, the volume, bib-

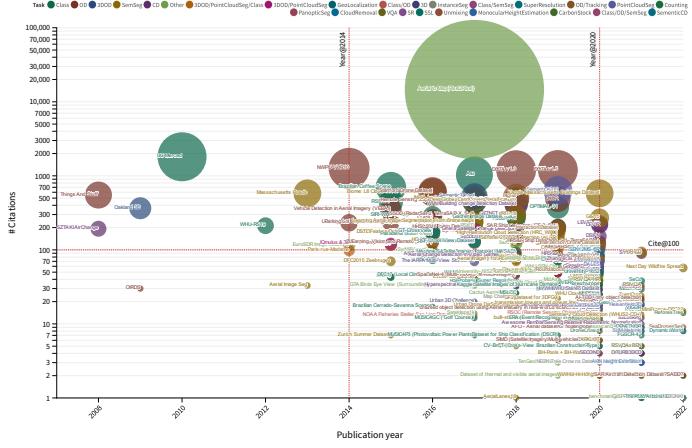


Fig. 2. Chronological overview of the number of citations of existing RS datasets. UC Merd [23], AID [24] and DOTA [25] are the most cited works. (Best viewed with zoom in)

liometric analysis, resolution distributions, research domains and dataset relationship are considered for a comprehensive dataset analysis.

- 3) To our knowledge, we are the first to measure and rank existing RS datasets using the dataset attributes provided in this study. Based on the ranking and selection, a new benchmark is built for the evaluation of RS methods.
- 4) We propose to analyze the relationships between different datasets. The dataset correlation matrix provides a new perspective to the RS community for the exploration of new algorithms across different datasets.
- 5) We release an open platform, termed EarthNets, for EO tasks. EarthNets aims to enable fair comparisons, efficient development of methods, and push the RS data to a larger research community.

The rest of this paper is organized as follows. Section 2 reviews the existing RS datasets for different tasks. Section 3 presents the analyses of the reviewed datasets from five different perspectives. Section 4 introduces the proposed dataset ranking and selection method, as well as the benchmark building process. Section 5 describes the newly released EarthNets platform. Section 6 presents the benchmarking results and analysis on the five selected RS datasets. Section 7 concludes the paper.

2 REMOTE SENSING DATASET REVIEW

In this section, we review and organize existing 401 public RS datasets into five parts regarding their tasks. Four of them are common tasks with lots of RS datasets, including image classification, object detection, semantic segmentation, and change detection. The fifth part contains lots of other RS tasks, including visual question answering (VQA) [26], [27], weather forecasting [28], 3D reconstruction [29], [30], [31], [32], Geo-localization [33], [34], [35], air quality [36], cloud removal [37], [38], [39], unmixing [40], image captioning [41], [42] and so on.

To analyze deeper on these datasets, we collect as much detailed information as possible for each dataset. Compared with existing review articles, our work provides richer information on the attributes of datasets. Specifically, the following ten aspects are considered.

Research domain. RS datasets are constructed for applications of different domains. In order to provide a clear presentation of these reviewed datasets, we organize them according to the specific domains in which they are created. Some typical research domains include agriculture, building, road, cloud, land use land cover (LULC), general scenes or objects, and so on.

Publication year. We provide the year of publication of the dataset, which is useful for chronological analysis of these datasets.

Number of samples: To measure the dataset scale, we list the number of samples for each dataset. Note that it could be the number of images (for image-based datasets), the number of video clips (for video-based datasets), the number of points (for point cloud datasets), or the number of image pairs (for change detection datasets).

Size of sample: The size of each sample in the dataset is also an important factor for measuring the dataset scale.

Volume: The volume of RS datasets should also be factored in as a measurement of the dataset scale.

Number of Classes: For image classification, object detection and semantic segmentation tasks, we provide the number of classes annotated in each dataset.

Data modality: For different EO tasks, a wide range of imaging sensors can be used to build datasets. The data sources of RS datasets may be but not limited to optical image, multi-spectral image, hyperspectral image [43], SAR [44], point cloud [45], and DSM (digital surface model) [46], [47].

Resolution range: The spatial resolutions of RS images have high correlation with the image content. It is also highly relevant to specific EO tasks to which it can be applied.

Number of citations: We provide the number of citations for each dataset to measure its popularity in the RS community.

Dataset link: To facilitate the research, we provide the download link of each dataset. More detailed information can be found on <https://earthnets.nicepage.io/>.

2.1 RS Image Classification Datasets

Image classification is a fundamental task in both the CV and RS communities. With image- or patch-level annotations, RS image classification has been involved in many different real-world applications. In Table 1, 91 RS image classification datasets are reviewed and presented. In order to facilitate researchers to search and index, we organize them into different research domains. To be specific, Table 1 contains 13 agriculture-related datasets [48], [49], [50], [51]. For agriculture-related applications, the images or patches in the datasets are labeled with binary labels (crop/non-crop) or crop type labels (up to 348 granular labels). For general scene classification, 16 datasets are presented in the table. Among them Million-AID is the largest one, which contains a million instances for training and evaluation of

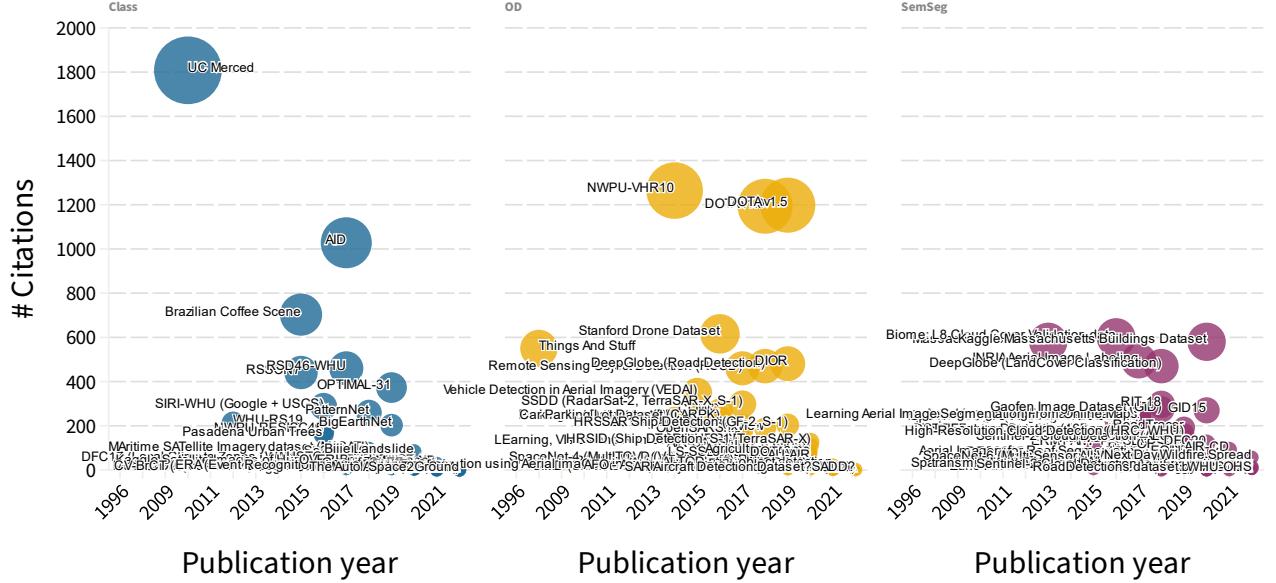


Fig. 3. Visualization of the citation information of three different tasks. This figure shows that scene recognition and object detection datasets have more citations than segmentation datasets. However, object detection and semantic segmentation datasets attract more research attention after year 2019.

scene classification methods. MLRSNet [52], RSD46-WHU [53], and NWPU-RESISC45 [54] are annotated with more than 40 class labels. There are 19 datasets for LULC applications in the table. Among these datasets, multiple types of data sources are used, including hyperspectral [43], multispectral [19], [55], [56], [57], SAR [58] and RGB data [59], [60].

There are 5 ship-related [61], [62], [63], [64] and 5 flood-related [65], [66], [67] datasets for RS image classification task. For cloud-related research domain, 4 datasets are re-reviewed in this table. Some specific domains with fewer datasets are also presented, like smoke [68], sea lion [69], solar power plants [70] and wind datasets [71]. It is worth noting that species classification datasets are annotated with the most semantic labels.

Compared with object detection (with object-level annotation) and pixel-level segmentation tasks, agriculture-related applications are mainly modeled as image classification tasks. In contrast, aircraft [44] and ship-related datasets are built primarily for object detection tasks. For agriculture-related datasets, data from the Sentinel satellites is mostly used with a lower spatial resolution to cover larger crop areas. General scene classification and LULC are the two dominant domains in RS image classification datasets. From the bibliometric view, general scene classification and LULC datasets are cited much more than other research domains. This indicates that scene classification is a heavily studied research direction in the RS community.

2.2 RS Object Detection Datasets

Object detection has close relationship with real-world applications like autonomous driving, video Surveillance and many other high-level scene understanding tasks. Thus, lots of popular works have been published in the CV community, like Faster RCNN [124], SSD [125], YOLO [126], and Transformer-based detectors [127]. For RS object detection

task, more and more, larger and larger datasets are also published for different EO applications, including aircraft detection [128], [129], [130], building detection [131], [132], [133], [134], ship detection [135], [136], [137], [138], vehicle detection [139], [140], [141], [142], [143], general ground object detection [25], [144], [145], [146], [147], [148] and other research domains [149], [150], [151].

In Table 2, 91 RS object detection datasets are reviewed and organized into 18 different research domains. The most popular domains for RS object detection tasks are general object detection with 12 datasets, building detection with 12 datasets, aircraft detection with 8 datasets, ship detection with 14 datasets and vehicle-related object detection with 20 datasets.

Similar to RS image classification tasks, datasets for general object detection are with higher citation numbers. Since object detection is a task with object-level annotations, the spatial resolutions of these datasets are usually higher than image classification datasets. However, for objects with large sizes, like ships, the data from Sentinel-1 and Sentinel-2 satellites is also used [152], [153]. For the detection of traffic objects [154] or other small objects [155], images captured from Unmanned aerial vehicle (UAV) are usually used.

2.3 RS Semantic Segmentation Datasets

Pixel-level semantic segmentation aims to assign semantic labels to each pixel of the input image. Compared with image-level and object-level tasks, interpreting the image data with semantic maps can provide a more complete understanding of the scene. In Table 3, we review and present 101 datasets for RS semantic segmentation tasks. Specifically, among them, 25 datasets are built for LULC or general scene segmentation tasks [21], [46], [47], [156], [157], [158], [159], [160], [161]. There are 18 datasets constructed for the segmentation of general scenes. There are 13 datasets [132], [170], [171], [172], [173], [174], [175] that

TABLE 1

Detailed information of 91 RS image classification datasets. These datasets are grouped into 27 different research domains in alphabetical order. Note that / denotes the missing information and the number of citations are retrieved as of Sep. 2022. The download links for all these datasets can be found at <https://earthnets.nicepage.io>.

Domain	Name	Year	#Samples	Sample Size	#Classes	Modality	Resolution	Vol.(GB)	#Cita.
Agriculture	Brazilian Coffee Scene [48]	2015	2876	64	2	RGB	20m	0.004	703
	Indian Pines [72]	2015	1	145	16	Hyperspectral	20m	0.0059	/
	Salinas [43]	2015	1	365	16	Hyperspectral	3.7m	0.026	/
	Crop Type Mapping Ghana [73]	2019	/	/	18	Sentinel-1,Sentinel-2,Planet	3~10m	312.54	/
	CV4A Kenya [74]	2020	4688	2016x3035	7	Sentinel-2	10m	3.5	9
	BreizhCrops [75]	2020	610000	/	9	Sentinel-2,MT	10~60m	/	34
	CaneSat [76]	2020	1627	10	2	Sentinel-2,MT	10m	0.006	/
	CropHarvest [77]	2021	90,480	12 ts	343	Sentinel-1,Sentinel-2,ERA5,DEM	10~60m	20	5
	South Africa Crop Type [78]	2021	122736	/	9	Sentinel-1,Sentinel-2	/	82.77	/
	DENETHOR [49]	2021	/	/	9	Sentinel-1,Sentinel-2	3m	254.5	9
Aircraft	The Canadian Cropland [51]	2022	78536	64	10	Sentinel-2	10m	26	0
Cloud	Space2Ground [79]	2022	10102	260	2	Sentinel-1,Sentinel-2,RGB	10~60m	0.501	0
	Sen4AgriNet [50]	2021	225000	366	158	Sentinel-2	10~60m	10240	1
	SARAC [44]	2022	4322	64	20	SAR	/	/	1
Event	SPARCS [80]	2016	80	1000	7	MS,Landsat	30m	1.43	168
	Kaggle Cloud Detection [81]	2019	9244	1750	4	RGB	/	5.86	0
	CloudCast [82]	2020	70080	1229	10	NWP	3km	320.31	6
Flood	Sentinel-2 Cloud Mask Catalogue [83]	2020	513	1022	3	Sentinel-2	20m	15.38	/
	ERA [84]	2020	343680	640	25	RGB Video	/	6.3	12
Forest	Hurricane Damage [65]	2019	16000	128	2	RGB	1m	0.064	33
	SEN-12-FLOOD [66]	2020	336	512	2	RGB,SAR,MS	10m	12.2	14
	Sen1Floods11 [67]	2020	4831	512	1	Sentinel-1	10m	14.3	40
	OMBRIA [85]	2022	3376	256	2	Sentinel-1,Sentinel-2	10m~20m	0.19	2
	FloodNet [13]	2021	2343	4000	9	RGB	0.015m	2.1	31
General Scenes	Kaggle Planet Forest [86]	2017	150000	256	17	RGB-NIR	5m	32.23	/
	OVERHEAD MNIST [87]	2020	1000	28	9	Grayscale	/	0.017	35
	fMoW [20]	2018	523846	/	63	RGB,MS	0.3m	3500	146
	UC Merced [23]	2010	2100	256	21	RGB	0.3m	0.3	1808
	WHU-RS19 [88]	2012	1013	600	19	RGB	0.5m	0.1	212
	RSSCN7 [89]	2015	2800	400	7	RGB	/	0.086	441
	NWPU-RESISC45 [54]	2016	31500	256	45	RGB	0.2~30m	0.404	176
	RSC11 [90]	2016	1232	512	11	RGB	/	0.63	75
	AID [24]	2017	10000	600	30	RGB	3m	2.4	1028
	RSD46-WHU [53]	2017	117000	256	46	RGB	0.5~2m	11	460
Multi-label Scenes	PatternNet [91]	2018	30,400	256	38	RGB	0.062~4.693m	1.3	260
	OPTIMAL-31 [92]	2019	1860	256	31	RGB	/	0.024	373
	MLRSNet [52]	2020	109,161	256	46	RGB	0.1~10m	1.254	25
	CLRS [93]	2020	15000	256	25	RGB	0.26~8.85m	1.735	16
	Million AID [14]	2021	1,000,000	150~550	28	RGB	0.5~153m	133.5	29
	NaSC-TG2 [94]	2021	20000	256	10	RGB-NIR	100m	/	8
	Satellite Image Classification [95]	2021	5631	256	4	RGB	/	0.023	/
	MultiScene [96]	2021	100,000	512	36	RGB	0.3~0.6m	0.85	1
	Hephaestus [97]	2022	216106	224	6	InSAR	/	93.71	0
	MUSIC4GC (Golf Course) [98]	2017	83431	16	2	MS,Landsat	30m	0.37	12
Hot Area	MUSIC4HA (Hot Area) [98]	2022	2511	16	6	Sentinel-2	10m	0.01	1
	Iceberg	Iceberg Detection [99]	2018	10028	75	SAR	/	0.295	/
Land Cover	SAT-4 [100]	2015	500000	28	4	RGB-NIR	1~6m	7.25	43
	SAT-6 [100]	2015	405000	28	6	RGB-NIR	1m	5.65	43
	Botswana [43]	2015	1	875	14	Hyperspectral	30m	0.077	/
	TiSeLaC [101]	2017	23	2866x2633	9	RGB-NIR,MT	30m	/	/
	Gaofen Image Dataset (GID) [102]	2018	150	7200	15	RGB-NIR	4m	71.1	274
	MSLCC [58]	2018	2	5596x6031,8149x5957	4	SAR,MS	10m	0.5	26
	BigEarthNet [55]	2019	590326	120	43	Sentinel-1,Sentinel-2	10m, 20m, 60m	121	203
	Slovenia Land Cover [103]	2019	940	500	10	Sentinel-2	10m	11.55	/
	So2sat LCZ42 [19]	2019	400673	32	17	Sentinel-1,Sentinel-2	10m	50.59	75
	TG1HRSSC [104]	2021	204	512	9	Hyperspectral	5m, 10m, 20m	0.277	4
Land Use	SIRI-WHU (Google+USGS) [59]	2016	2400	200	12	RGB	2m	0.7	288
	RSI-CB256 [60]	2017	24000	256	35	RGB	0.3~3m	2.2	37
	RSI-CB128 [60]	2017	36000	128	45	RGB	0.3~3m	0.88	37
	Austin Zoning [105]	2017	3,666	773x961	5	RGB	/	0.596	/
	HistAerial [106]	2019	42000	25,50,100	7	Grayscale	/	7.6	29
	AiRound [56]	2020	11753	300	11	RGB,Sentinel-2,Ground,Aerial	/	33	5
	CV-BrCT [56]	2020	24000	500	9	RGB	/	9.2	5
Landslide	EuroSAT [57]	2018	27000	64	10	Sentinel-2	10m	1.92	32
	SenseEarth classify [107]	2020	70000	100~12655	51	RGB	0.2~153m	10.8	/
	Bijie Landslide [108]	2020	2773	200	2	RGB	0.68m	0.51	84
Military	MSTAR-Sclass [109]	1996	9466	368	8	SAR	0.3m	0.444	/
Plant/Tree	Forest Cover Type [110]	2013	581012	12	7	Tree Attributes	/	0.07	/
	Pasadena Urban Trees	2016	100,000	/	18	RGB	/	/	161
	Aerial Cactus Identification [111]	2019	17000	32	2	RGB	/	0.025	/
	WiDS Datathon 2019 [112]	2019	11000	256	2	RGB	3m	0.46	/
	The Auto Arborist Dataset [113]	2022	2,637,208	1,024	344	RGB,MS	/	24	1
	TreeSatAI [114]	2022	50381	304x304,6x6	47	Sentinel-1,Sentinel-2,RGB	10m,0.2m	16.3	0
Sea lion	Forest Damages Larch Casebearer [115]	2021	1543	1500	5	RGB	UAV	3.3	/
	NOAA Sea Lion Population Count [69]	2017	950	4900	4	RGB	/	96	13
Ship	Ships in Satellite Imagery [61]	2017	4000	80	2	RGB	3m	0.343	/
	MASATI [62]	2018	7389	512	7	RGB	0.08~2m	2.3	93
	DSCR [63]	2019	20,675	150~800	7	RGB	/	/	7
	FGSCR-42 [63]	2021	9320	140~800	42	RGB	/	4.76	7
	SynthWakeSAR [64]	2022	46080	96000	10	SAR	3.3m	4.3	0
Smoke	USTC_SmokeRS [68]	2019	6225	256	6	RGB	1000m	0.79	49
Solar Power Plants	MUSIC4P3 [70]	2017	1280000	16	2	MS,Landsat	30m	4.6	7
Species	GeoLifeCLEF 2021 [116]	2021	19,000,000	256	31,435	RGB-IR,MS,LC,DEM	1m,0.3m,0.1m	840	19
Tailings Dam	BrazilIDAM [117]	2020	769	384	2	RGB	10~60m	57	11
Urban Village	S2UC [118]	2021	1714	224	2	RGB	2m	1.8	1
Vegetation	Kennedy Space Center [119]	2015	1	550	13	Hyperspectral	0.18m	0.055	/
	Brazilian Cerrado-Savanna [120]	2016	1311	64	4	MS	5m	0.011	17
Vehicles	WAMI DIRSIG [121]	2017	55226	64	2	Hyperspectral	0.3m	0.33	48
	Kaggle Find a Car Park [122]	2019	3262	1296	2	RGB	/	2.75	/
	MAFAT-Fine-Grained [123]	2021	4216	/	37	RGB	0.05~0.15m	/	/
Wind	Airbus Wind Turbine Patches [71]	2021	155,000	128	2	RGB,MS	1.5m	1	/

TABLE 2

Detailed information of 91 RS object detection datasets. These datasets are grouped into 18 different research domains in alphabetical order. Note that / denotes the missing information and the number of citations are retrieved as of Sep. 2022. The download links for all these datasets can be found at <https://earthnets.nicepage.io>.

Domain	Name	Year	# samples	Size	# classes	Modality	Resolution	Volume (GB)	# Citations	
Aircraft	PASTIS [162]	2021	2433	128	18	Sentinel-2	10m	29	15	
	PlanesNet [128]	2017	32000	20	2	RGB	/	0.4	/	
	MTARS1 (Aircraft) [163]	2019	9385	256	2	RGB	/	0.48	/	
	RarePlanes [164]	2020	713348	512	110	MS,WorldView3	0.3~1.5m	310.55	32	
	CGI Planes [165]	2021	500	/	2	RGB	/	0.7	/	
	CASIA-aircraft [130]	2021	58,121	399	2	RGB	/	/	/	
	Airbus Aircraft Detection [129]	2021	109	2560	2	RGB	0.5m	0.092	/	
Bridge	SAR Aircraft [166]	2022	2966	224	2	SAR	0.5m~3m	0.18	2	
	Military Aircraft [167]	2022	3842	800	20	RGB	/	1.1	0	
	Bridges Dataset [168]	2019	500	4800x2843	2	RGB	0.5m	1.45	6	
Building	SpaceNet-4 (Multi-View) [134]	2018	60000	900	1	MS,WorldView2	0.3m	186	45	
	DeepGlobe (Building) [131]	2018	24586	650	2	Panchromatic,RGB,MS	0.5m	/	470	
	WHU Building [169]	2018	25577	512	2	RGB	0.3m	24.41	414	
	CrowdAI Mapping [170]	2018	401,755	300	1	RGB	/	5.3	/	
	Map Challenge [170]	2018	341,058	300	2	RGB	/	/	23	
	TBF [171]	2018	13	40,000	2	RGB	/	/	/	
	Microsoft Building (Australia) [172]	2019	11,334,866	/	2	RGB	/	6.4	/	
General	Microsoft Building (Uganda/Tanzania) [173]	2019	17,942,345	/	2	RGB	/	3.5	/	
	Microsoft Building (USA) [132]	2019	129,591,852	/	2	RGB	/	34.4	/	
	Microsoft Building (Canada) [174]	2019	11,842,186	/	2	RGB	/	2.5	/	
	Urban Building Classification [175]	2022	800	600	61	RGB	0.5~0.8m	0.675	/	
	BONAI [176]	2022	3,300	1,024	1	RGB	0.3m~0.6m	4.86	1	
	NWPU-VHRR10 [144]	2014	800	1000	10	RGB,IRRG	0.08~2m	0.07	1264	
	RSOD [145]	2017	976	1000	4	RGB	0.3~3m	0.077	460	
General	TGRS HRSSD [146]	2017	21761	10569	13	RGB	0.15~1.2 m	8.6	126	
	xView [147]	2018	1,413	3,000	60	RGB,MS	0.3m	20	188	
	fMoW [148]	2018	523846	/	63	RGB,MS	0.3m	3500	146	
	DOTA v1.0 [25]	2018	2806	4000	15	RGB	/	18	1193	
	DOTA v1.5 [177]	2019	2806	4000	16	RGB	/	18	1198	
	DIOR [15]	2019	23463	800	20	RGB	0.5~30 m	6.93	481	
	DOTA v2.0 [178]	2020	11268	4000	18	RGB	0.1~0.81	34.3	69	
Human/Animals	iSAID [179]	2020	2806	4000	15	RGB	/	18	110	
	VALID [180]	2020	6690	1024	30	RGBD	/	15.7	15	
	UAVOD10 [149]	2022	844	1000~4800	10	RGB	0.15m	0.9	0	
	BIRDSSAI [181]	2020	162000	640	10	Thermal	UAV@60-120m	3.7	25	
	Land Covers	Dstl Satellite Imagery [150]	2017	57	3,348	10	RGB,MS	0.3m~7.5m	21.7	/
	Object Counting	RSOC (Object Counting) [182]	2020	3057	2500	4	RGB	/	0.082	14
	Oil and Gas Tank (OGST) [183]	2020	10000	512	2	RGB	0.3m	1.87	/	
Oil Storage Tanks	Airbus Oil Storage Detection [151]	2021	103	2560	2	MS	1.2m	0.102	/	
	Oil Storage Tanks [184]	2019	10000	512	2	RGB	0.5m	3	/	
	Volcanoes	Hephaestus [185]	2022	216106	224	InSAR	/	93.71	0	
	Person	Semantic Drone-OD [186]	2019	400	5000	2	RGB	/	3.91	/
	Stanford Drone [187]	2016	100	1400x1904	6	RGB Video	0.025m	69	616	
	Sea	NOAA Sea Lion Count [69]	2017	950	4,900	4	RGB	/	96	13
	Aerial Maritime Drone [188]	2020	508	800x600	5	RGB	/	0.038	18	
Search/Rescue	SeaDronesSee [189]	2022	5630	3,840~5,456	6	RGB	/	60.3	9	
	AFO-Floating objects [190]	2020	3647	720~3840	6	RGB	/	4.7	9	
	Search And Rescue [191]	2021	2552	1000	1	RGB	0.5m	/	3	
	Ship	OpenSARShip [135]	2017	11346	900	1	Sentinel-1	10m	1.7	176
	SSDD [136]	2017	1160	500	2	SAR	1~15m	/	298	
	HRSC2016 (Ship) [192]	2017	1061	300x300~1500x900	26	RGB	0.4~2m	3.74	203	
	Kaggle Airbus Ship [137]	2018	192556	768	2	RGB	1.5m	31.41	/	
Small Objects	Airbus Ship Detection [137]	2018	40,000	768	2	RGB	/	31.4	/	
	Ships in Google Earth [138]	2018	794	2000	2	RGB	/	2	/	
	SAR Ship Detection [193]	2019	43819	256	2	SAR	3m, 5m, 8m,10m	0.4	204	
	AIR-SARShip-1.0 [194]	2019	31	3000	2	SAR	1~3m	0.24	0	
	AIR-SARShip-2.0 (GF-3) [161]	2020	300	1,000	2	SAR	1~3m	0.22	108	
	LS-SSDD (Large Scale) [152]	2020	15	20,000	2	Sentinel-1,SAR	0.5,1,3m	7.8	69	
	HRSID (Ship) [153]	2020	5,604	800	2	Sentinel-1,SAR	0.5~3m	0.58	128	
Traffic Objects	SWIM-Ship [195]	2021	14610	768	2	RGB	0.5~2.5m	12.5	0	
	CASIA-Ship [196]	2021	1,118	1,680	2	RGB	/	/	/	
	xView3-SAR [197]	2022	1000	~29400x24400	2	Sentinel-1	10m	1500	0	
	Tree	AI-TOD [154]	2021	28036	800	8	RGB	0.3m~30m	42	22
	SODA-A [155]	2022	2510	4761x2777	9	RGB	/	/	0	
	AU-AIR [198]	2020	32823	1920	8	RGB	UAV@30m	2.2	57	
	HighD [199]	2018	110000	4096x2160	2	RGB	/	/	519	
Vehicles	Interaction Dataset [200]	2019	10,933	/	1	RGB	/	/	196	
	Intersection Drone [201]	2020	11,500	4096x2160	5	RGB	/	/	120	
	Roundabouts Drone [202]	2020	13,746	4096x2160	8	RGB	/	/	52	
	NEON Tree Crowns [203]	2020	11,000	100 million trees	2	RGB	/	27.4	3	
	Forest Damages [204]	2021	1543	1500	5	RGB	UAV	3.3	/	
	Things And Stuff (TAS) [139]	2008	30	792	2	RGB	0.5m	0.01	550	
	OIRDS [205]	2009	900	256~640	5	RGB	0.15m	0.153	30	
Vehicles	UCAS_AOD [206]	2014	976	1,000	2	RGB	/	3.24	232	
	VEDAI(Vehicle) [140]	2015	1,250	1,024	9	IRGB	0.125m	3.9	350	
	PKLott(Parking Lot) [207]	2015	12417	1280	2	RGB	UAV	4.6	236	
	COWC [141]	2016	388435	256	2	RGB	0.15m	62.5	265	
	Car Parking Lot (CARPK) [208]	2016	1448	1280	2	RGB	UAV	2	235	
	DLR3k/DLR-MVDA [142]	2016	20	3744	7	RGB	0.13m	0.162	/	
	ITCV (Vehicle) [143]	2018	173	5616	2	RGB	0.1m	12	47	
Vehicles	VisDrone2019-DET [209]	2019	10209	2000x1500	10	RGB	UAV	2	48	
	VisDrone2019-VID [209]	2019	40000	3840x2160	5	RGB	UAV	14	48	
	VisDrone2019-SOT [209]	2019	139300	3840x2160	3	RGB Video	UAV	68	48	
	VisDrone2019-MOT [209]	2019	40000	3840x2160	5	RGB Video	UAV	14	48	
	SIMD (Multi-vehicles) [210]	2020	5000	1024	15	RGB	UAV@150m	1	6	
	VisDrone [211]	2020	275,437	1,400	11	RGB	UAV	16	46	
	EAGLE [212]	2020	8820	936	2	RGB	0.05~0.45m	/	16	
Swimming pool/car	MOR-UAV [213]	2020	10948	1080	1	RGB	UAV	/	21	
	DroneVehicle [214]	2021	56,878	840	5	RGB-Infrared	UAV@100m	13.09	4	
	Swimming pool/car [215]	2019	3750	224	2	RGB	/	0.12	/	
	ArtifiVe-Potsdam [216]	2021	4800	600	1	MS	0.05m	15.6	2	

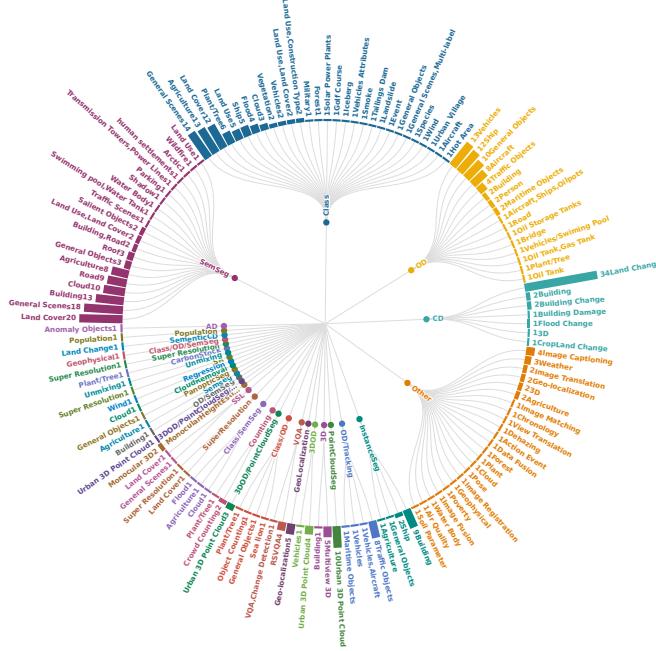


Fig. 4. Visualization of the relationships between tasks and research domains. It can be seen that there are strong correlations between research domain and EO tasks.

are designed for building extraction with pixel-level annotations. Note that, some of them are constructed for the building instance segmentation tasks. In this case, the buildings are annotated with both the object-level and pixel-level labels. Datasets designed for LULC, general scenes, building, and road segmentation are dominating the RS semantic segmentation tasks. For road extraction, 9 datasets [16], [217], [218], [219], [220], [221], [222] are constructed. For the cloud-related applications, there 10 datasets built with lower spatial resolution RS images than other domains [223], [224], [225], [226], [227]. Furthermore, 9 agriculture datasets are annotated with pixel-level labels [162], [228], [229], [230], [231]. From the bibliometric view, building and road extraction and highly cited domains. This makes sense because building and road segmentation tasks are widely used in real-world applications.

2.4 RS Change Detection and Other Tasks

RS change detection aims to quantitatively analyze the surface changes based on remotely-sensed data. It is critical for real-world applications like damage monitoring and urban planning. In Table 4, we present 35 datasets that are built for RS change detection tasks. Most of them are focusing on the change of land cover or lance use. Several other datasets are constructed for 3D change detection [232], building-specific change detection [233], and flood-related change detection [234]. Many different data modalities are used for constructing these datasets including optical (RGB), point cloud, hyperspectral, multi-spectral and SAR. Note that some datasets are annotated with pixel-wise semantic labels, like SECOND [235] and Dynamic World [236] not only the change/non-change binary label.

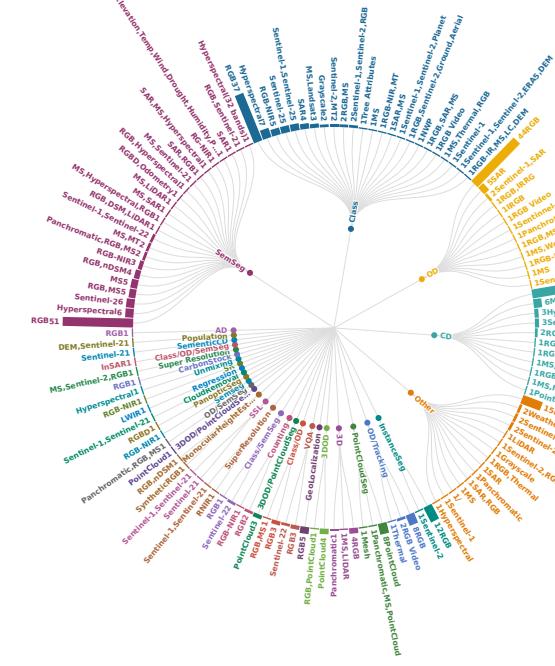


Fig. 5. Visualization of the relationships between data modalities and different RS tasks. Although there is a wide range of image sources used for EO, optical data (RGB) is still the most used modality for the majority of RS tasks.

Apart from the RS image classification, object detection, semantic segmentation and change detection tasks, we also list 80 datasets constructed for other tasks. For example, image captioning datasets [41], [42] and visual question answering datasets [26], [27] are combine RS data with natural languages. Multi-view stereo datasets [237], [238], [239] for 3D reconstruction. There are also some sporadic tasks like geo-localization [33], [34], weather forecasting [28], [240], oil parameter estimation [241] and wind speed estimation [71], [242] datasets.

3 REMOTE SENSING DATASET ANALYSIS

In this section, we will analyze the reviewed more than 400 RS datasets and provide the statistics from five different aspects.

3.1 The volume Trend

Thanks to the powerful representation learning capabilities, deep learning networks trained with large-scale datasets have shown superior performance than classical machine learning methods. In the deep learning era, large-scale datasets play an important role in training deep models with better performance as well as better generalizability. Another advantage of large-scale datasets is that it aligns better with real-world scenarios. In Fig. 1, we visualize the chronological overview of the volumes for 401 RS datasets. Note that the volume (GB) shown in this figure is transformed into the logarithmic scale, and larger circles indicate larger volumes. It can be clearly seen that datasets before the year 2015 are usually smaller in volume. Similar to the CV community, after deep learning became the mainstream technique, both the number and the volume of RS datasets

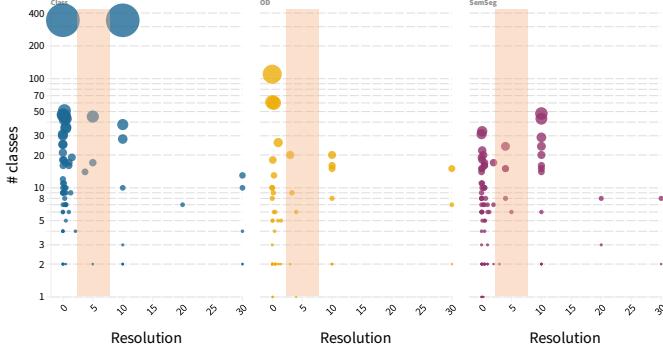


Fig. 6. Visualization of the relationships between data resolution and the number of annotated classes. An interesting finding is that most datasets are with the resolution of lower than 1m or larger than 10m. The datasets with resolution in the range of between 1 to 10m is obviously scarce.

significantly increase. For example, the volumes of fMoW [148] and Sen4AgriNet [50] are larger than 4000 GB. Well-annotated large scale datasets can greatly help the RS community in developing and evaluating more powerful deep learning models with better performance and generalizability.

3.2 The Bibliometric Analysis

We show the citation information in a chronological order in Fig. 2. The number of citations are collected as of Sep. 2022. From the figure we can see that the dataset with the highest number of citations is Aerial2map [243]. Aerial2map is a dataset for image translation, and the pixel2pixel algorithm is the reason of its high number of citations. In general, Fig. 2 shows that datasets published during the year range of 2014 to 2020 have a higher number of citations.

UC Merced [23], AID [24] for scene classification, NWPU VHR10 [54], DOTA [25], [177] for object detection are datasets with high number of citations. After the year of 2020, we can see that VQA dataset [26] and SeCo dataset [244] for self-supervised learning have attracted more and more research attention in the RS community.

In Fig. 3, we display the citation information of three different tasks, namely, the RS image classification, object detection and semantic segmentation. According to this figure, we can clearly see that scene recognition and object detection datasets have more citations than segmentation datasets. However, there are more object detection and semantic segmentation datasets published after the year of 2019. This indicates that object-level and pixel-level understanding of RS images are becoming more and more popular in the RS community.

3.3 Analysis on data modalities

To analyze the image sources used in the RS community, we summarize and visualize the data modalities used for different RS tasks. The relationships between data modalities and RS tasks are shown in Fig. 5. Although there is a wide range of image sources, optical data (RGB) is still the most used modality for the majority of RS tasks. In Fig 4, we display the relationships between tasks and research domains.

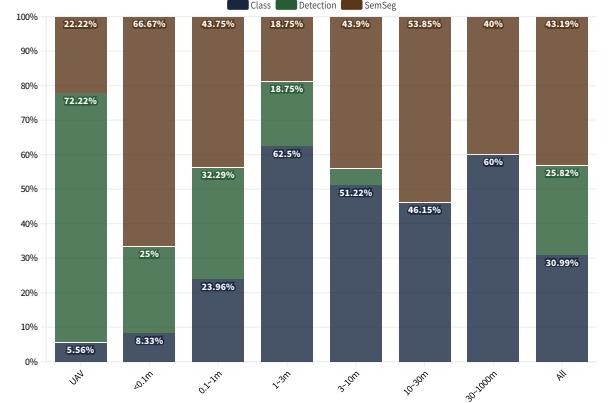


Fig. 7. Visualization of the task distributions with regard to different spatial resolutions. Datasets with UAV images are mostly constructed for object detection. Datasets with the resolution range of 1m to 30m are mostly designed for image classification and semantic segmentation tasks.

3.4 Analysis on spatial resolutions

For RS images, the spatial resolution has a high correlation to the image content. In Fig. 6, we show the relationships between data resolution and the number of annotated classes. In general, this figure clearly shows that the number of semantic classes for RS image classification tasks is higher than for RS object detection and segmentation tasks. The reason is that object-level and pixel-level datasets require much more annotation efforts when the number of semantic classes increases.

Another interesting finding is that most datasets are with a resolution of lower than 1m or larger than 10m. The datasets with resolution in the range of between 1 to 10m are obviously scarce. The reason for this phenomenon is that many EO applications require either high-resolution (<1m) imagery or global coverage (Sentinel 1&2, >10m). However, the EO data with resolution 1~10m also has great potential in a range of applications. More research attention should be paid to fill in this gap.

The task proportion distribution with regard to different spatial resolutions is displayed in Fig. 7. As shown in this figure, datasets with UAV images are mostly constructed for object detection. More than 60% of the datasets with very high-resolution (<0.1m) are designed for RS semantic segmentation task. For datasets with a resolution range of 1m to 30m, RS image classification and semantic segmentation take the largest proportion. Only less than 20 percent of the datasets are built for RS object detection in this resolution range.

3.5 The correlation between different datasets

To provide a more global view of the 401 RS datasets, for the first time, we propose to analyze the correlation between different datasets based on the provided attribute information in this study.

We treat the attribute information of each dataset as a data sample, and measure the similarity between different datasets. Formally, let n represent the number of samples, s represent the size of each sample in dataset D . We denote v as the volume of D . Then the scale of D can be

TABLE 3

Detailed information of 101 RS semantic segmentation datasets. These datasets are grouped into 19 different research domains in alphabetical order. Note that / denotes the missing information and the number of citations are retrieved as of Sep. 2022. The download links for all these datasets can be found at <https://earthnets.nicepage.io>.

Domain	Name	Year	#Samples	Sample Size	#Classes	Modality	Resolution	Vol.(GB)	#Cita.
Agriculture	Agricultural Crop Cover [228]	2018	40	/	2	MS	30m	4.4	
	GF2 Dataset for 3DFGC [229]	2019	101	2,652	5	RGB-NIR	4m	0.056	22
	TimeSen2Crop [230]	2020	1000000	10980	16	Sentinel-2	10m	1.1	11
	Agriculture-Vision [231]	2020	94986	512	9	RGB-NIR	0.1~0.2m	4.4	81
	WHU-Hi-LongKou [245]	2020	1	550x400	9	Hyperspectral	0.463m	/	2
	WHU-Hi-HanChuan [245]	2020	1	1217x303	16	Hyperspectral	0.109m	/	2
	WHU-Hi-HongHu [245]	2020	1	940x475	22	Hyperspectral	0.043m	/	2
Arctic	ZueriCrop [246]	2021	28000	24	48	Sentinel-2	10m	39	24
	EuroCrops [247]	2021	805,401	0.5 ha	43	Sentinel-2	10m	8.6	2
Building	Arctic Sea Ice Image Masking [248]	2021	3392	357x306	8	RG-NIR	10m	0.092	/
	SpaceNet-1 (Building) [249]	2016	9735	650	2	RGB,MS	0.5~1m	31	231
	SpaceNet-2 (Building) [249]	2017	24586	650	2	RGB,MS	0.3m	104	231
	INRIA Aerial Image Labeling [218]	2017	360	1500	1	RGB	0.3m	19.5	494
	built-structure-count dataset [250]	2019	5364	512	1	RGB	0.3m	2.23	12
	SpaceNet-6 (Multi-Sensor All Weather) [251]	2020	3401	900	2	SAR,RGB	0.5m	55.9	51
	SpaceNet-7 (Multi-Temporal Urban) [252]	2020	1525	1024	2	MS,MT	4m	20.1	18
	SynthInet-1 [253]	2020	2108	572	2	RGB	0.3m	0.977	39
	Kaggle buildings segmentation [133]	2020	6038	256	2	RGB	/	0.899	/
	Kaggle Massachusetts Buildings [254]	2020	151	1500	2	RGB	1m	2.93	580
	Open Cities AI Challenge [255]	2020	11,000	1,024	2	RGB	0.03~0.2m	81.5	/
Cloud	Mini Inria Aerial Image Labeling Dataset [218]	2021	32,500	512	2	RGB	0.3m	/	/
	High-speed Rail Line Building Dataset [256]	2021	336	2000	2	RGB	0.5m	/	/
	AIS From Online Maps [257]	2017	1671	3000	2	RGB	0.5m	23.8	243
	Biome: L8 Cloud Cover [223]	2016	96	/	4	RGB	30m	96	599
General Objects	38-Cloud [224]	2018	17601	384	2	RGB	30m	13	73
	Sentinel-2 Cloud Detection (ALCD) [258]	2019	38	1830	2	MS,Sentinel-2	10~60m	0.234	141
	HRC_WHU [226]	2019	150	1280x720	2	RGB	0.5~15m	0.17	165
	WHU Cloud Dataset [39]	2020	859	512	2	RGB	30m	3.56	25
	95-Cloud [225]	2020	34701	384	2	RGB	30m	18	12
	WHUS2-CD+ [227]	2021	36	10980	2	Sentinel-2	10m	27.8	13
	AIR-CD [259]	2021	34	7300	2	RGB-NIR	4m	13	93
	The Azavea Cloud Dataset [260]	2021	32	/	2	Sentinel-2	10m~60m	/	/
	Sentinel-2 Cloud Cover [261]	2022	22728	/	2	MS	10m~60m	51.2	/
	DLRSD [262]	2018	2100	256	17	RGB	0.3m	0.004	70
General Scenes	Kaggle aerial segmentation [160]	2020	72	800	6	RGB	/	0.033	/
	AIR-PoSAR-Seg [161]	2022	2000	512	6	SAR	8m	0.609	/
	ISPRS 2D - Potsdam [46]	2011	38	6000	6	RGB,nDSM	0.05m	15.625	/
	ISPRS 2D - Vaihingen [47]	2011	33	2200	6	RGB,nDSM	0.09m	16.6	/
	Aerial Image Segmentation [263]	2013	80	512	2	RGB	0.3~1m	0.007	33
	DFC2015 Zeebruges [264]	2015	7	100,000	8	RGB,DSM,LiDAR	0.05m	0.0024	70
	Zurich Summer Dataset [265]	2015	20	1000	8	RGB-NIR	0.61m	0.38	7
	DSTL Feature Detection (3Band) [266]	2016	450	3391	10	RGB	0.31m	13.82	174
	DSTL Feature Detection (16Band) [266]	2016	1350	3391	10	MS	1.24m,7.5m	7.84	174
	EvLab-SS Dataset [267]	2017	60	4500	11	RGB	0.1m,0.25m	/	42
General Scenes	SynthAer [268]	2018	765	1280	8	RGB	/	0.977	/
	Aeroscapes [269]	2018	3269	1280	11	RGB	UAV@5-50m	0.73	48
	Urban Drone Dataset (UDD) [270]	2018	301	4,096	6	RGB	UAV	1.1	18
	RIT-18 [157]	2018	3	9393x5642,8833x6918,12446x7654	18	MS	0.047m	1.5	297
	Semantic Drone Dataset-SemSeg [186]	2019	400	5000	20	RGB	/	3.91	/
	DroneDeploy [271]	2019	55	6,000	7	RGB	0.1m	/	/
	Mid.Air [272]	2019	420000	1024	12	RGBD,Odometry	/	1000	38
	AeroRIT [273]	2019	1	3975x1973	6	RGB,Hyperspectral	0.4m	1.8	18
	SemCity Toulouse [274]	2020	16	3500	8	MS	0.5~2m	8.8	10
	UVAVid [275]	2020	420	4000	8	RGB	UAV	5.88	75
Settlements	DFC21-DSE [276]	2021	98	800	2	SAR,MS,Hyperspectral	10~750m	18	/
Land Cover	Washington DC MALL [156]	2013	1	1,280	7	Hyperspectral	/	0.14	/
	Pavia Center [43]	2011	1	1096	9	Hyperspectral	1.3m	0.121	/
	Pavia University [43]	2011	1	610	9	Hyperspectral	1.3m	0.032	/
	DeepGlobe (LandCover) [277]	2018	1146	2448	7	RGB	0.5m	2.96	470
	HyRank [278]	2018	5	1,000	14	Hyperspectral	30m	0.4	5
	WHDLDD [262]	2018	4940	256	6	RGB	2m	0.102	80
	SEN12MS [159]	2019	541,986	256	17	MS,SAR	10m	510	119
	Urban Semantic 3D (DFC19) [279]	2019	2783	1024	6	MS,LiDAR	0.3~1.3m	285	/
	XiongAn [280]	2019	1	3,750	19	Hyperspectral	0.5m	3	/
	Chesapeake Land Cover [281]	2019	100000	224	6	RGB,MS	1m	404.95	52
Land Use	DFC20 [282]	2020	180662	256	10	Sentinel-1,Sentinel-2	10m	9.6	121
	BDCI2020 [283]	2020	145,981	256	7	RGB	2m	1.3	/
	LandCoverAI [158]	2020	41	9000	3	RGB	0.25m,0.5m	1.4	44
	GID15 [102]	2020	150	6800x7200	15	RGB,MS	4m	18	270
	LoveDA [284]	2021	5,987	1,024	7	RGB	0.3m	9.6	21
	MiniFrance-DFC22 [285]	2022	2322	2000	15	RGB	0.5m	93	15
	GeoNRW [286]	2022	7783	1000	10	RGB,nDSM	1m	32	8
	SEASONET [21]	2022	1759830	120	33	Sentinel-2	10m	229	0
	TimeSpec4LULC [287]	2022	/	262 months	29	MS,MT	500m	60	0
	Five-Billion-Pixels [288]	2022	150	7200x6800	24	RGB,MS	4m	104	0
Parking	WHU-OHS [289]	2022	7795	512	24	Hyperspectral	10m	94.9	0
	OpenSentinelMap [290]	2022	137,045	192, 96	15	RGB,Sentinel-2	10m~60m	455	
	DFC18 [291]	2018	10,798	2,001	20	MS,Hyperspectral,RGB	0.05~1m	10.1	136
	MultiSenGE [292]	2022	8157	256	14	Sentinel-1,Sentinel-2	10m	530	0
Road	APKLLOT [293]	2020	501	/	2	RGB	/	3	6
Roof	Massachusetts Roads [254]	2013	1171	1500	1	RGB	1m	10.56	580
	ERM PAIW [220]	2015	41	4000	2	RGB	0.3m	0.635	117
	HD-Maps [219]	2016	20	4000	5	RGB	0.3m	0.146	133
	SpaceNet-3 (Road) [249]	2017	3711	1300	2	Panchromatic,RGB,MS	0.3~1.24m	106	231
	RoadNet [217]	2018	20	/	2	RGB	0.21m	0.905	89
	AerialLanes18 [294]	2018	20	5616	1	RGB	0.125m	0.0014	1
	SpaceNet-5 (Road Network) [249]	2019	2369	1300	2	Panchromatic,RGB,MS	0.3m	84	/
	SpaceNet-8 (Flooded Road) [295]	2022	/	1300	4	Panchromatic,RGB	0.3~0.8m	/	/
	RoadTrace [221]	2019	3,000	4,096	1	RGB	0.6m	/	192
	Microsoft RoadDetections [222]	2022	20000	1088	1	RGB	1m	9.25	0
Salient Objects	AIRS [296]	2019	1047	10000	1	RGB	0.075m	17.6	73
	Open AI Challenge: Caribbean [297]	2019	7	52,318	5	RGB	0.04m	/	/
	RID [298]	2022	2000	/	16	RGB	0.1m	1.5	0
Shadow	ORSSD [299]	2019	800	500	8	RGB	/	0.026	104
	EORSSD [299]	2020	2,000	500	2	RGB	/	0.06	74
Water Tank	AISD [300]	2020	514	512	2	RGB	/	0.29	25
Traffic Scenes	BH-Pools+WaterTanks [301]	2020	350	3000	2	RGB	/	1.9	4
Power	DLR-SkyScapes [302]	2019	16	4680	31	RGB	0.13m	/	52
Water Body	TIPL [303]	2020	1100	3840	3	RGB	UAV	4.2	19
Wildfire	Kaggle Water Bodies [304]	2020	2841	1000	2	RGB	/	0.28	/
	Next Day Wildfire Spread [305]	2022	18,445	64	2	Multi-source	1000m	4	57

TABLE 4

Detailed information of 35 RS change detection datasets. These datasets are grouped into 5 different research domains in alphabetical order. Note that / denotes the missing information and the number of citations are retrieved as of Sep. 2022. The download links for all these datasets can be found at <https://earthnets.nicepage.io>.

Domain	Name	Year	#Samples	Sample Size	#Classes	Modality	Resolution	Vol.(GB)	#Cita.
3D	URB3DCDD [232]	2021	50	/	2	PointCloud	0.5 pm	1.5	4
Building	AIST Building Change Detection [306]	2017	16950	160	2	RGB	0.4m	17.77	87
	WHU Building change detection [169]	2018	2	15354x32507	2	RGB	0.075m	5	409
	LEVIR-CD [307]	2020	637	1024	2	RGB	0.5m	2.64	224
	xView2 (xBD) [308]	2018	22068	1024	4	RGB	0.5m	51	80
CropLand	CropLand Change Detection (CLCD) [309]	2022	600	512	2	RGB	0.5~2 m	0.5	0
Flood	California flood dataset [234]	2019	1	1534x808	2	RGB,MS	5m,30m	0.33	50
Land Change	SZTAKI AirChange [310]	2008	13	800	2	RGB	1.5m	0.04	193
	Taizhou Data [311]	2014	1	400	4	MS	30m	/	/
	Kunshan Data [311]	2014	1	800	3	MS	30m	/	/
	Cross-sensor Bastrop [312]	2015	4	444x300,1534x808	2	MS	30m,120m	/	/
	GETNET dataset [313]	2018	1	463x241	2	Hyperspectral	30m	0.05	297
	Onera Satellite CD [314]	2018	24	600	2	Sentinel-2	10m	0.48	186
	AICD [315]	2018	1000	800	2	RGB	/	1.7	84
	CDD (season-varying) [316]	2018	16000	256	2	RGB	0.03~0.1m	2.7	149
	Hyperspectral CD [317]	2018	3	984x740,600x500,390x200	5	Hyperspectral	30m	1.7	33
	HRSCD [314]	2019	291	10000	5	RGB	0.5m	5	86
	MIS-WH [318]	2019	190	150	2	RGB-NIR	1m	0.43	/
	SECOND [235]	2020	4662	512	6	RGB	0.5~3m	2.2	4
	Zhang et al. CD dataset [319]	2020	4	1431x1431,458x559,1154x740	2	RGB,NIR	2m,2.4m,5.8m	0.1	74
	DSIFN [320]	2020	3,988	512	2	RGB	10m	0.46	152
	Hermiston City Oregon [321]	2018	1	390x200	5	Hyperspectral	30m	/	/
	Hi-UCD [322]	2020	1293	1024	9	RGB	0.1m	/	19
	Google Data Set [323]	2020	19	1000~5000	2	RGB	0.55m	0.6	57
	DFC21-MSD [276]	2021	2250	4000	15	MS,MT	1~30m	325	/
	Relative Radiometric Normalization [324]	2021	7	300~5000	2	MS	0.31m,0.4m,10m,20m,30m,60m	1	10
	HTCD [325]	2021	2	11 Kx15 K,1.38 Mx1.04 M	2	RGB	0.5971m, 0.074m	1.74	4
	S2Looking [233]	2021	5000	1,024	2	RGB	0.5~0.8m	10.21	12
	SYSU-CD [326]	2021	20,000	256	2	RGB	0.5m	5.17	89
	WH-MAVS [327]	2021	47,134	200	15	RGB	1.2m	/	/
	S2MTCP [328]	2021	1520	600	/	MS	10m	10.6	18
	Dynamic EarthNet Challenge [49]	2021	22500	1024	7	RGB	3m	/	0
	MSBC [329]	2022	3,769	256	2	RGB,SAR,MS	2m	3.9	0
	Dynamic World [236]	2022	/	/	9	Sentinel-2	10m	/	14
	MSOSCD [329]	2022	5,107	256	2	RGB,SAR,MS	10~60m	2.7	0

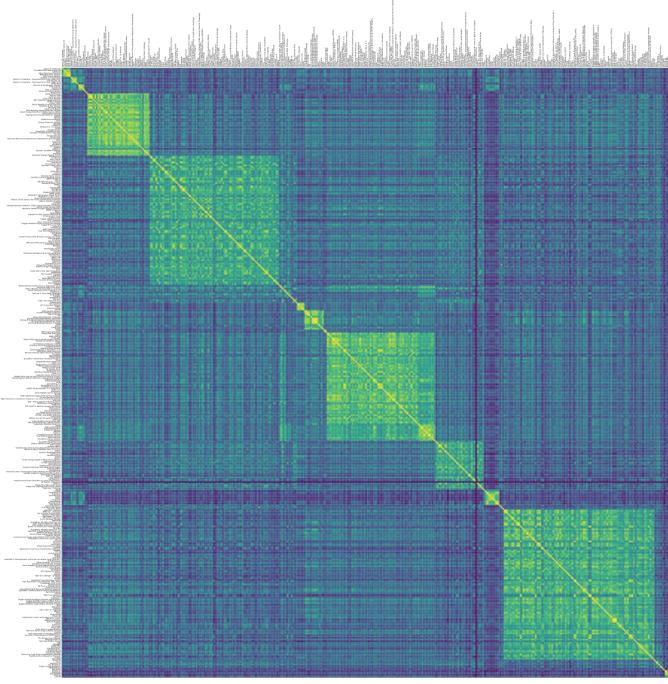


Fig. 8. Visualization of the correlation between different datasets. For the first time, we propose to analyze the correlations between different RS datasets based on their attribute information collected in this study (Best viewed in color).

quantitatively measured using n , s and v . Furthermore, we quantify the annotation level of D and represent it using m . Specifically, we assign 1 to m for image-level annotation, 2 for object-level, 3 for pixel-level, 4 for instance-level, 5 for panoptic-level, and 0 for no-label. Similarly, we also quan-

tify the task of D to t . According to the task type, t can be 1 for RS image classification, 2 for object-detection, 3 for semantic segmentation, 4 for change detection and 0 for other tasks. Then we use c to represent the number of annotated classes in dataset D , and r to denote the max resolution of samples in D . With these definitions, n, s, v, m, c, r, t are numerical values representing the attributes of D .

Since the research domain of a dataset is provided by a word or phrase, it is non-trivial to measure the distance between them. For example, the research domain "Ship" should be closer to "Sea" than "Tree" or "Aircraft". The research domain "Tree" should be more similar to "Forest" not "Building". To this end, we propose to use the pre-trained word embedding [380] models to compute the real-valued vector feature d for each research domain. Here we denote d as the textual features of the domain.

Following these pre-processing pipelines, we are able to quantify the attributes of dataset D into two feature vectors: one is $F = [n, s, v, m, c, r, t]$, the other is the word embeddings for the research domain d . For two datasets D_1 and D_2 , we use F_1 , F_2 and d_1 , d_2 to represent the features of these two datasets. Then, we can compute the similarity using the following formula

$$\begin{aligned} \cos(\theta_1) &= \frac{F_1 \cdot F_2}{\|F_1\|_2 \|F_2\|_2}, \\ \cos(\theta_2) &= \frac{d_1 \cdot d_2}{\|d_1\|_2 \|d_2\|_2}, \\ sim(D_1, D_2) &= \cos(\theta_1) + \cos(\theta_2). \end{aligned} \quad (1)$$

In Fig. 8, the correlation matrix between 401 datasets is visualized. The lighter the color, the higher the similarity. To our knowledge, this is the first work that analyze the correlation between all the RS datasets. The correlation matrix reflect the distances between different pair of the

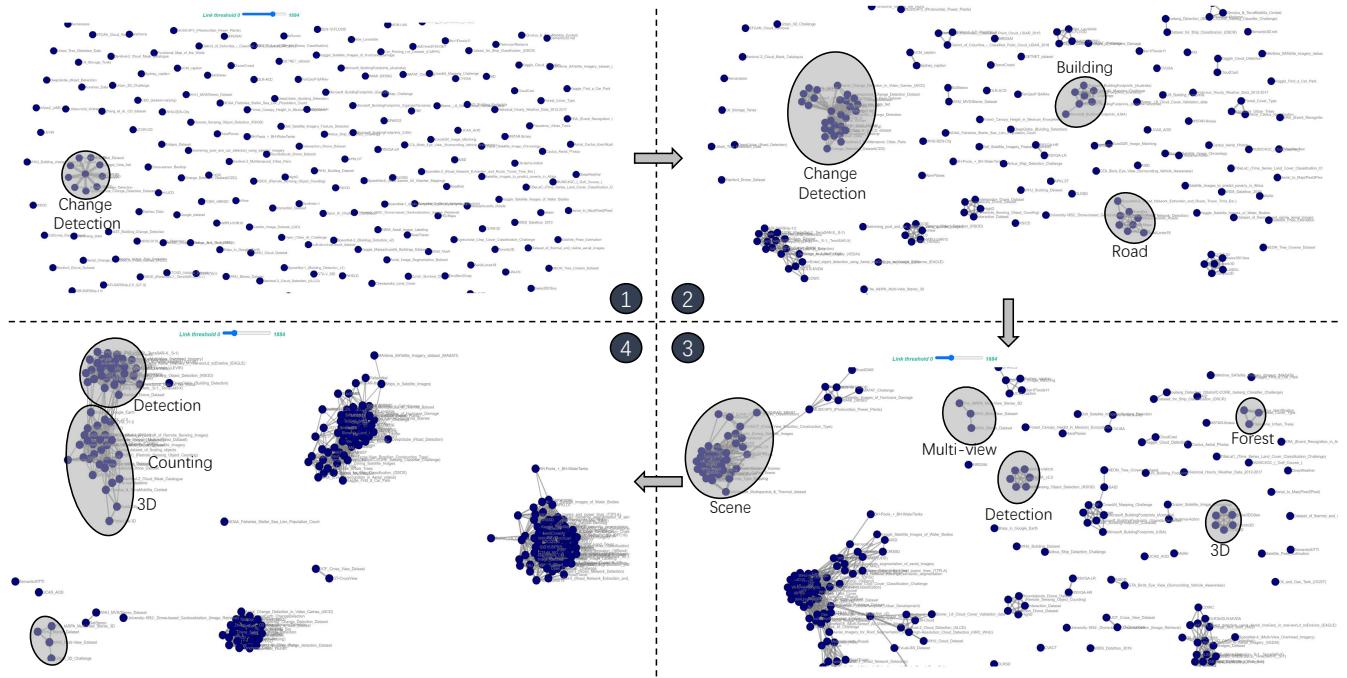


Fig. 9. With the correlation matrix, we can visualize the RS datasets using network graphs. The node represents the RS dataset, and the link between nodes denotes the similarity between them. Datasets gradually cluster together when the connecting threshold decreases.

RS dataset. The distance information could be valuable for the RS community. Here we provide some possible ways to leverage the correlation between RS datasets for future research.

- 1) **Dataset recommendation.** Based on the relationships between datasets, we are able to recommend similar and related datasets given one dataset. This is helpful for researchers to find desired datasets for their research tasks.
- 2) **Domain Adaptation.** Domain adaption aims to improve the performance of a model on a target domain using the knowledge learned in the source domain. With the correlation map, researchers can easily find the proper source and target datasets for developing novel domain adaptation algorithms.
- 3) **Dataset Assembling.** The distance between datasets can also be used to ensemble multiple small but similar datasets into a larger one for training large-scale deep networks.
- 4) **Multi-task model Training.** Similarly, using the distance between datasets, we can also combine datasets with similar spatial resolution, data modalities, research domains, but different tasks into a unified one for training multi-task deep models.

Furthermore, with the correlation matrix, we can visualize the RS datasets using an interactive network graph. The node represents the RS dataset, and the link between nodes denotes the similarity between them. In Fig. 9, we can see that different datasets gradually cluster together when the connecting threshold decreases.

4 DATASET RANKING AND BENCHMARK BUILDING

Researchers in the RS community have been publishing more and more datasets to benefit the development of new methods. However, algorithms can easily saturate their performance on these datasets [381]. Deep learning models can achieve almost perfect performance on small-scale or domain-specific datasets. However, small-scale datasets are easier to have bias and can not reflect the performance of methods in real-world complex scenarios [14]. Methods developed on small datasets or specific domains are difficult to generalize to other scenarios.

Considering these disadvantages, it is urgent to employ new benchmarks with large-scale, general research domain, and high annotation quality datasets for a fair and consistent evaluation of RS methods. Although the attributes of a large number of datasets are provided, it is still not intuitive to compare the quality of different datasets. Thus, for the first time, in this study, we propose to rank these datasets based on their attributes.

4.1 Dataset Ranking Metrics

Regarding the desirable properties of benchmark datasets, long et al. [14] propose the DiRS formula, namely, the diversity, richness, and scalability of datasets. These properties are good references for designing metrics to measure and rank the RS dataset. However, it is non-trivial to quantitatively measure the diversity and richness of existing datasets. In order to approximate DiRS metric, we consider both the data diversity and annotation diversity in this study.

To measure the data diversity, we first examine the research domain of one dataset. Some datasets constructed with specific domains will have limitations on the diversity

TABLE 5

Detailed information of 83 other RS datasets. These datasets are grouped into 33 different research domains in alphabetical order. The download links for all these datasets can be found at <https://earthnets.nicepage.io>.

Domain	Name	Year	#Samples	Sample Size	#Classes	Modality	Resolution	Vol.(GB)	#Cita.
Action Event	Okutama-Action [330]	2017	77000	3840x2160	12	RGB	UAV@10~45m	25.9	130
Agriculture	Paddy Rice Maps South Korea [331]	2022	12942	256	/	Sentinel-1	10m	0.198	/
	Paddy Rice Labeling South Korea [332]	2022	/	/	/	Sentinel-2	/	0.0016	/
Air Quality	Air Quality e-Reporting [36]	2021	/	/	/	/	/	/	/
Anomaly Objects	Aerial Anomaly Detection [333]	2022	/	/	2	RGB	UAV	/	0
Building	Urban 3D Challenge [334]	2017	157,000	/	2	RGB	0.5m	/	20
Chronology	Draper Satellite Image Chronology [335]	2016	1,000	3100x2329	/	RGB	/	36	
Cloud	STGAN Cloud Removal [37]	2019	217190	256	2	Sentinel-2,RGBIR	10m	1.5	35
	SEN12MS-CR [38]	2020	122218	256	/	Sentinel-1,Sentinel-2	10~60m	272	99
Counting	DLR-ACD [336]	2019	33	4458	2	RGB	0.045~0.15m	/	28
	DroneCrowd [209]	2020	3360	1920x1080	2	RGB	UAV	1	8
	RSOC [182]	2020	3057	2500	4	RGB	/	0.082	14
Data Fusion	SEN1-2 [337]	2018	282384	256	/	Sentinel-1,Sentinel-2	10m	42.68	122
	QXS-SAROPT [338]	2021	40000	256	/	SAR,RGB	1m	2.7	9
Dehazing	SateFazelk [339]	2017	1,200	512	/	RGB	3m	1.2	14
Forest	Forest Canopy Height [340]	2018	1105	/	/	LIDAR	1m	62.9	0
Self-supervised	SeCo [244]	2021	24000x5	264	/	Sentinel-2	10m	43.6	38
	SSL4EO-S12 [341]	2022	1004316	264	/	Sentinel-1,Sentinel-2	10m~60m	1500	0
Geo-localization	CVUSA [33]	2017	44,416	750x750	0	RGB	0.3m	15.7	131
	University-1652 [34]	2020	146,580	/	/	RGB	/	/	49
	UCF Cross View Dataset [35]	2017	35404	1200x1200	0	RGB	/	59.5	130
	CVACT [342]	2019	128334	1200x1200	0	RGB	0.5m	152.8	71
	University-1652 [34]	2020	50220	1024x1024	0	RGB	/	58.4	49
	VIGOR [343]	2021	90618	640x640	0	RGB	0.114m	94.2	16
Geophysical	TenGeoP-SARwv [344]	2019	37000	/	10	SAR	5m	31.7	3
	UCM caption [41]	2016	2,100	256	21	RGB	0.3m	0.255	97
Image Captioning	Sydney caption [41]	2016	613	500	7	RGB	0.5m	0.298	97
	RSICD [41]	2017	10,921	224	30	RGB	/	0.46	210
	LEVIR-CC [42]	2022	10077	512	/	RGB	/	/	/
Image Matching	EuroSDR Image Matching [345]	2014	/	/	/	RGB	6~13m	/	116
Image Registration	Dataset of thermal and visible [346]	2019	110	336,4000	0	RGB,Thermal	/	0.31	2
Image Translation	Aerial to Map(Pixel2Pixel) [243]	2017	2,194	600	/	RGB	/	0.239	14625
	WHU-SEN-City [347]	2019	18542	256	/	Sentinel-1,Sentinel-2	10m	4.3	57
Multiview 3D	The IARPA Multi-View Stereo 3D [348]	2017	/	/	/	MS,LIDAR	0.3m	75	70
	SatStereos [237]	2019	144	/	/	Panchromatic	0.5m	127	10
	WHU MVS/Stereo Dataset [238]	2020	1776	5376	/	RGB	0.1m	95.7	29
	WHU TCL SatMVS [349]	2021	300	5,120	/	Panchromatic	2.1~2.5m	29	2
	WHU Multi-View Dataset [349]	2020	28400	768	/	RGB	0.1m	12.3	29
	WHU Stereo Dataset [349]	2020	21868	768	/	RGB	0.1m	8.4	29
Plant/Tree	Cactus Aerial Photos [350]	2018	24,000	32	2	RGB	/	0.053	26
	Reforestree [351]	2022	105	4000	/	RGB	0.02m	7.5	14
	Urban Tree Detection Data [352]	2020	60	256	2	RGB-NIR	0.6m	0.114	0
Population	Sc2Sat Population [353]	2022	/	/	/	DEM,Sentinel-2	10m	/	0
Pose	Satellite Pose Estimation [354]	2019	15303	1920x1200	/	Grayscale	/	4.6	77
Poverty	Poverty in Africa [355]	2020	32823	256	/	RGB	/	5.58	/
RSVQA	RSVQA-LR [26]	2020	772	256	9	Sentinel-2	10m	0.15	38
	RSVQA-HR [356]	2020	10659	512	89	RGB	0.15m	14.4	38
	RSVQAXBEN [26]	2021	590,326	up to 120	26,875	Sentinel-2	10m	28.6	5
	RSIVQA [26]	2021	37,264	256~4000	864	RGB	0.1~8m	29.5	33
Soil Parameter	Hyperview Challenge [241]	2022	2,886	11~250	/	Hyperspectral	2m	1.85	0
	ALSAT-2B [357]	2021	5518	256	/	RGB-NIR	2.5m/10m	0.044	0
Super Resolution	The WorldStrat [358]	2022	3449	1054	8	MS,Sentinel-2,RGB	1.5~60m	109.5	2
	SEN2VENimus [359]	2022	132955	256	/	Sentinel-1,Sentinel-2	10m~20m	87	/
	Proba-V Super Resolution [360]	2018	1160	384	/	RNR	100~300m	0.71	37
Unmixing	DLR HySU [40]	2021	1	86x123	/	Hyperspectral	0.3m~1m	0.005	2
Urban 3D Point Cloud	CORE3D [361]	2018	/	/	/	Panchromatic,MS,PointCloud	/	/	0
	benchmark_ISPRS2021 [362]	2021	20	1024	/	RGB	0.08m	3.22	1
	SemanticKITTI [363]	2019	4549M	39.2 km ²	25	PointCloud	/	80.2	632
	Toronto3D [29]	2020	78.3M	1.0 km ²	8	PointCloud	/	1.1	69
	Swiss3DCities [30], [31]	2020	226M	2.7 km ²	5	PointCloud	/	/	17
	DALES [364]	2020	505.3M	10.0 km ²	8	PointCloud	/	/	61
	LASDU [365]	2020	3.12M	1.02 km ²	5	PointCloud	Aircraft @ 1200m	/	23
	Campus3D [32]	2020	937.1M	1.58 km ²	14	PointCloud	UAV	/	18
	SensatUrban [366]	2020	2847.1M	6 km ²	13	PointCloud	UAV	20.6	51
	Heisenheim3D [239]	2021	125.7M	0.19 km ²	11	PointCloud	/	58.8	18
	SUM-Helsinki [367]	2021	19M	4 km ²	6	Mesh	0.075m	8.35	8
	Oakland 3-D PointCloud [368]	2009	1.6 M	1.5 km	5	PointCloud	/	0.033	365
	District of Columbia LiDAR 2015 [45]	2015	/	/	8	PointCloud	/	34.4	/
	Semantic3D.net [369]	2017	4000 M	/	8	PointCloud	/	24.5	526
	District of Columbia LiDAR 2018 [45]	2018	/	/	11	PointCloud	/	279.8	/
	Paris-Lille-3D [370]	2017	143 M	1.94 km ²	50	PointCloud	/	19.9	183
	DublinCity [371]	2019	260M	2.0 km ²	13	PointCloud	/	166.8	49
	Paris-rue-Madame [372]	2014	20 M	1 km	26	PointCloud	/	/	98
Vehicles	AM3D-Real [373]	2022	1,012	720x480	2	RGB,PointCloud	0.25m	/	0
VQA,Change Detection	CDVQA [27]	2022	2968	512	19	RGB	0.5~3m	/	0
Water Body	Forel-Ule Index global inland waters [374]	2021	/	/	/	MS	500m	0.004	/
Weather	Historical Hourly Weather [375]	2017	/	/	/	Weather Attributes	/	0.075	/
	DeepWeather [28]	2020	20	/	9	Weather Attributes	/	2,965	58
	ClimateNet [240]	2020	459	16 (768,1152)	9	NC variables	/	28	22
	EarthNet2021 [376]	2021	32000	128	/	Sentinel-2	20m	614.4	7
Climate	Earth Surface Temperature [377]	2017	16	/	/	Climate variables	/	0.09	/
	Greenhouse Gas [378]	2017	/	/	/	Gas Emission	/	0.0001	/
	Hurricane Wind Speed [379]	2020	114634	366	/	LWIR	/	2.24	12

of data sources. Hence, we first filter out them and only keep datasets designed for general purposes, like LULC or general scene understanding. Next, we choose to measure the dataset scale using the number of samples and the volume of the dataset, i.e., attribute variables n, s, v . Furthermore, we take the modality diversity, namely, the number of data modalities k into consideration. Since images with higher spatial resolution can provide richer visual content, we also factor the resolution r in as a part of the metric.

Considering the annotation richness, we use the number of annotation classes c and the quantified annotation level m to measure the richness of the labels.

To start with, given 401 RS datasets, we first filter out datasets designed for specific domains. After this step, 114 datasets remain as candidates. Then, we use the aforementioned attributes to quantitatively measure the diversity and richness of each dataset. Since there exist significantly high values in the n, s, v of different datasets, we use log normalization to standardize them into the range 0 to 1.

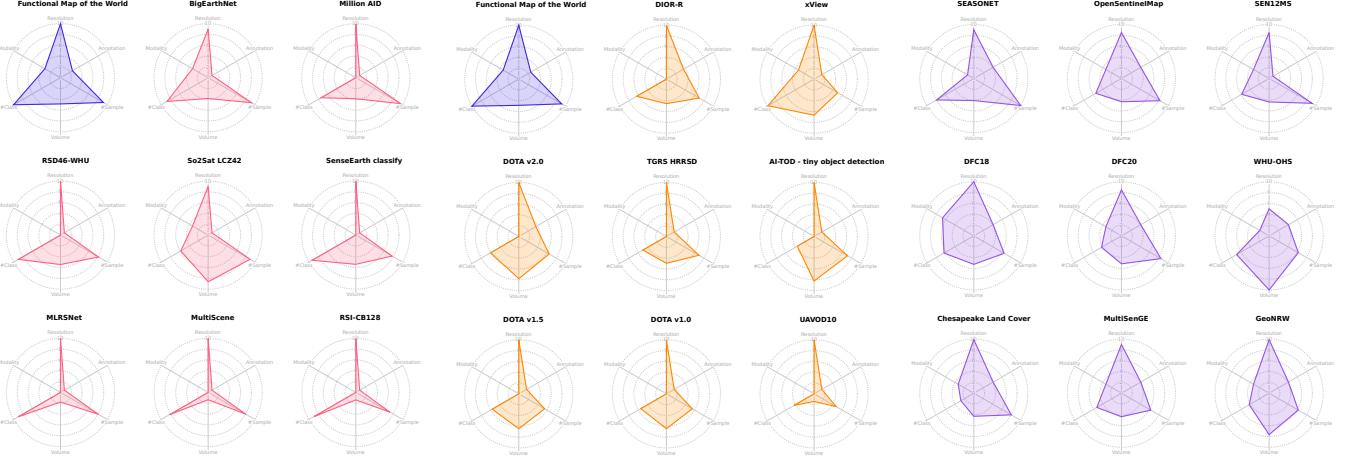


Fig. 10. Radar charts visualization of some top ranked RS datasets for three different tasks. six different attributes are compared and displayed in this figure. Note that all the attribute values are normalized to the range of 0 to 10.

Next, we normalize each of the dataset attributes in r, m, c, k into the range 0 to 1. Finally, we add them together to form the final score for the given dataset.

Based on the measurement defined above, we can compute the scores and rank these RS datasets. Based on the rankings, we select datasets for three different tasks. Specifically, for the RS image classification task, the top five ranked datasets are: 1). fMoW [148], 2) BigEarthNet [55], 3) Million AID [14], 4) So2Sat LCZ42 [19], 5) RSD46-WHU [53]. As for the RS object detection task, the top five datasets are 1) fMoW [148], 2) DIOR [15], 3) xView [147], 4) DOTA v2.0 [178], 5) TGRS HRRSD [146]. When it comes to RS semantic segmentation task, the top five ranked datasets are 1) SEASONET [21], 2) OpenSentinelMap [290], 3) SEN12MS [159], 4) GeoNRW [286], 5) Five-Billion-Pixels [288]. In Fig. 10, we use the radar charts to compare the attributes of some top ranked datasets. A whole list of the charts is displayed on <https://earthnets.nicepage.io>.

4.2 Dataset Selection for Benchmarking

We aim to select several datasets designed with general purpose, large diversity and high richness for developing and evaluating deep learning methods. Although there are lots of large-scale RS datasets that meet these standards, it is unacceptable and not environment-friendly to benchmark all of them for the evaluation of RS algorithms. Thus, in this study, we choose to select two datasets for each task, including one with high-resolution and one with low resolution for larger geographical coverage.

Following this instruction, the following datasets are selected. 1) **fMoW** with high resolution ($\sim 1m$) data and **BigEarthNet** with low resolution ($>10m$) imagery are selected for image classification. 2) **DIOR** with high-resolution ($\sim 1m$) data and **fMoW** with large objects are selected for the RS object detection task. 3) **GeoNRW** with high-resolution ($\sim 1m$) images and **SEASONET** with low-resolution ($>10m$) images are selected for RS semantic segmentation. To sum up, there are five datasets selected to build a unified benchmark for three different tasks.

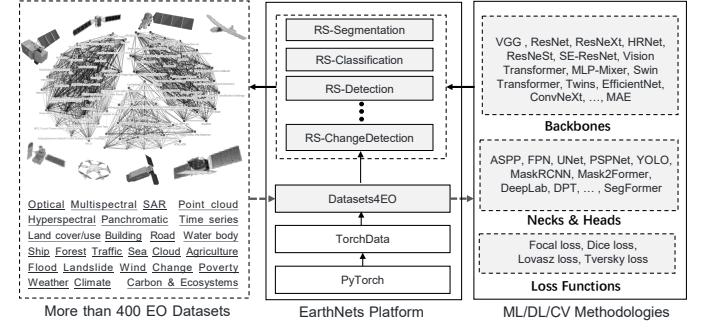


Fig. 11. The architecture design of the proposed EarthNets platform. Basically, EarthNets is based on PyTorch [382] and TorchData. It contains the Dataset4EO for a standard and easy-to-use dataset loading library, and some high-level libraries for different EO tasks.

5 THE EARTHNETS OPEN PLATFORM

Large-scale and high-quality datasets are important for a faithful evaluation of RS algorithms. While other factors like training tricks, hyper-parameters, optimizers, and initialization methods are also quite critical for a fair and reliable comparison of different methods. Thus, an open platform is crucial for the fair evaluation, reproducibility, and efficient development of novel methods. However, there is still no unified deep learning platform for different RS tasks. Torchgeo [22] mainly focuses on the data loading part. AiTLAS [381] mainly contains codebase for RS classification task. In contrast, we aim to build a new unified platform for RS community, which not only deals with dataset loading, but also includes libraries for different RS tasks.

5.1 Design

Fig. 11 illustrates the overall architecture design of the proposed EarthNets platform. Basically, we base the platform on PyTorch [382] and TorchData. The library Dataset4EO is designed as a standard and easy-to-use dataset loading library. Note that Dataset4EO can be used alone or together with our high-level libraries like RS-Classification, RS-Detection, and so on.

For the design of the EarthNets platform, we mainly consider two factors. The first one is the decoupling between dataset loading and high-level EO tasks. As we present in this study, there are more than 400 RS datasets with different file formats, data modalities, research domains and download links. Building a standard and scalable dataset loading library can largely accelerate the research for the whole RS community. Furthermore, researchers from other machine learning community can also benefit from the standard dataset loading library.

Another thing we consider is to push the RS data to a larger machine learning community. There are lots of novel deep learning models published in the CV and machine learning community, including different backbones, models and loss functions. The EarthNets platform is designed to easily apply these models to RS datasets, and so as to fill in the gap between RS and CV communities.

5.2 Implementation

In this section, we will detail the implementation of the main libraries in the EarthNets platform **Dataset4EO**. There are some difficulties in loading RS datasets, especially for researchers in other communities. 1) The datasets are with different downloading links, folder structures, file formats, data modalities, research domains and annotation levels. It will be helpful to download, decompress, and split the dataset automatically. 2) Images in RS datasets are usually with multiple modalities and bands. To accelerate the training speed, it is better to move the data augmentation to GPU. 3) For datasets with too large volumes or time series streaming data, support iterable-style data pipes will be useful to handle the dataset loading process.

Remote Sensing Tasks Based on the Dataset4EO library, we build the RS Classification, RS Detection and RS Segmentation libraries. All these libraries share the same dataset loading module. To establish a deep connection with the CV community, we base these libraries for RS tasks on the libraries from OpenMMLab [383]. As shown in Fig. 11, EarthNets enables an easy adaptation of modern deep learning models from CV community to RS community. For example, backbone models like ResNet [384], EfficientNet [385], ConvNext [386], Vision Transformers [387], MLP-Mixer [388], Swin Transformer [389], and so forth. Numerous state-of-the-art architectures designed for CV tasks like RetinaNet [390], UNet [391], Deeplab [392], YOLO [126], Upernet [393], and SegFormer [394] can be applied to RS data. By this means, EarthNets can serve as a bridge between the CV and RS communities.

6 EXPERIMENTS

In this section, we benchmark state-of-the-art deep learning models from the CV community on five selected RS datasets. We also compare them with the methods specifically designed for the RS datasets.

6.1 Benchmark Datasets

Five datasets are selected to benchmark the model performance using the EarthNets platform, including 1) **fMoW**, 2) **BigEarthNet**, 3) **SEASONET**, 4) **GeoNRW**, and 5) **DIOR**.

TABLE 6

Image classification results on the fMoW [148] dataset. Top-1 accuracy, precision, recall and F1 score are reported. The best results are in bold.

Methods	Image Classification				
	Pre-trained	Top-1	P	R	F1
ViT-Small	Random	54.1	53.45	51.8	52.03
	Random	43.11	40.33	41.09	40.18
MLP-Mixer	ImageNet	58.25	58.73	57	57.28
	ImageNet	58.8	58.7	57.01	57.33
ResNet-50	ImageNet	62.05	63.81	60.34	61.19
	ImageNet	66.42	66.33	65.29	65.5
EfficientNet-b4	ImageNet	62.05	63.81	60.34	61.19
	ImageNet	66.42	66.33	65.29	65.5
ConvNext-Small	ImageNet	62.05	63.81	60.34	61.19
	ImageNet	66.42	66.33	65.29	65.5
Swin-Tiny	ImageNet	62.05	63.81	60.34	61.19
	ImageNet	66.42	66.33	65.29	65.5

TABLE 7

Multi-label image classification results on the BigEarthNet [55] dataset. mAP, micro precision, micro recall and F1 score are reported. The best results are in bold.

Methods	Image Classification				
	Pre-trained	mAP	P	R	F1
ResNet-18* [244]	Random	79.80	-	-	-
	ImageNet	85.90	-	-	-
ResNet-18* [244]	MoCo-v2	85.23	-	-	-
	ImageNet	86.74	-	-	-
ResNet-50*	ImageNet	85.74	76.87	75.89	76.38
	ImageNet	84.48	73.84	77.19	75.48
EfficientNet-b4	ImageNet	85.59	73.64	79.91	76.65
	ImageNet	87.19	78.01	80.22	79.1
ConvNext-Small	ImageNet	82.76	73	74.36	73.67
	ImageNet	82.76	73	74.36	73.67
Swin-Tiny	ImageNet	82.76	73	74.36	73.67
	ImageNet	82.76	73	74.36	73.67
MLP Mixer	ImageNet	82.76	73	74.36	73.67
	ImageNet	82.76	73	74.36	73.67

6.2 Implementation Details

Optimizer: For convolution-based models, SGD is used as the optimizer. The AdamW [395] is used for optimizing the Transformer-based models. **Initialization:** By default, we use the ImageNet pre-trained weights for initializing the models. For some architectures with no ImageNet [18] pre-trained weights, we train them from scratch.

Other hyper-parameters including batch size and learning rate are set differently for each dataset. More implementation details are provided in the public codes at <https://github.com/EarthNets>.

Metrics: For multi-label image classification datasets, we report the following metrics: precision (P), recall (R), F1 score, and mean average precision(mAP). For precision, recall and F1, we set the threshold value to 0.5 for all models. As for object detection, mAP is used as the measurement for performance evaluation. Three metrics are used to evaluate the semantic segmentation task, including overall (micro-averaged) Accuracy (aAcc), mean (macro-averaged) Accuracy(mAcc), and mean Intersection over Union (mIoU).

6.3 Benchmarking Results and Comparisons

In this section, we benchmark the selected five datasets using the proposed EarthNets platform. In order to avoid huge amount of computation cost, we choose to evaluate some representative state-of-the-art (SOTA) methods from the CV community on the large-scale RS datasets.

Comparisons on the fMoW Dataset. fMoW [148] is a large-scale dataset built for recognizing the functional purpose of buildings and land use. It contains 1 million images from over 200 countries, annotated with 63 different classes. In this study, we use the fMoW-rgb version of the dataset for model evaluation. Table 6 reports the benchmarking results.

TABLE 8

Object detection results on the DIOR [15] dataset. The mAP performance is reported. The best results are in bold.

Method	Object Detection			
	Backbone	Optimizer	Epochs	mAP
RetinaNet* [15]	ResNet-50	-	-	65.7
RetinaNet* [15]	ResNet-101	-	-	66.1
PANet* [15]	ResNet-50	-	-	63.8
PANet* [15]	ResNet-101	-	-	66.1
Mask-RCNN* [15]	ResNet-50	-	-	63.5
Mask-RCNN* [15]	ResNet-101	-	-	65.2
YoloV3* [15]	DarkNet53	-	-	57.1
YoloV3	DarkNet53	SGD	120	64.0
YoloV3	Swin-Tiny	AdamW	120	64.6
YoloV3	ConvNext-Small	AdamW	120	67.6
Mask-RCNN	ResNet-50	SGD	120	68.5
Mask-RCNN	Swin-Tiny	AdamW	120	70.5
Mask-RCNN	ConvNext-Small	AdamW	120	72.4

TABLE 9

Semantic segmentation results on the SEASONET [21] dataset. We report the aAcc, mAcc and mIoU metrics. The best results are in bold.

Method	Semantic Segmentation				
	Backbone	Optimizer	Iter.	aAcc	mAcc
DeeplabV3* [21]	DenseNet121	-	-	-	47.53
DeeplabV3,PT* [21]	DenseNet121	-	-	-	48.69
DeeplabV3	ResNet-50	SGD	80k	82.87	58.49
DeeplabV3	ResNet-50	SGD	160k	83.52	62.65
DeeplabV3	ConvNext-Small	AdamW	120	81.36	56.31
DeeplabV3	Swin-Tiny	AdamW	120	82.75	61.5
Upernet	ResNet-50	SGD	120	83.2	60.36
SegFormer	MiT	AdamW	120	83.75	64.25
					53.87

In general, we can see that using the ImageNet pre-trained weights can largely improve the performance. When we compare the CNN-based methods with the Transformer-based one, we can find that Swin-Tiny [389] clearly outperforms other CNN-based methods regarding all the four metrics. Among the CNN-based methods, ConvNext [386] is the best performing one.

Comparisons on the BigEarthNet Dataset. BigEarthNet is a large-scale multi-label Sentinel-2 benchmark dataset. It consists of 590,326 Sentinel-2 image patches, annotated with the CORINE Land Cover classes. There are two versions of the labels, one with 43 categories and another with 19 categories. In this study, we adopt the new class nomenclature (19 categories) introduced in [397]. Regarding the evaluated methods, four CNN-based architectures (ResNet-18, ResNet-50, EfficientNet-b4, ConvNext). For transformer

based method, we evaluate the Swin-Tiny, which is usually missed in existing benchmarking results. Furthermore, MLP-based method, the MLP-Mixer [388] is also compared. Additionally, we also compare the results reported by existing work [244] on the BigEarthNet dataset.

Table 7 reports the benchmarking results. In general, the results indicate that Swin-Tiny performs best on this multi-label classification dataset. However, there is no significant advantage compared with other CNN-based methods. Another conclusion we can make is that ResNet-50 is a strong baseline method. From the results, it can be seen that ResNet-50, pre-trained on ImageNet or using self-supervised MoCo-V2 [398], can perform better than MLP-Mixer, EfficientNet-b4 on this dataset. The performance of ConvNext is competitive to ResNet-50, but lower than the transformer-based method Swin-Tiny. Note that * indicates that the results of the method are reported in existing work. Generally speaking, the results benchmarked using the EarthNets platform are higher than or competitive to existing reported results.

Comparisons on the DIOR Dataset. Table 8 presents the benchmarking results on the DIOR dataset built for the object detection task. We choose two representative and widely-used object detection methods designed from the CV community. To be specific, YoloV3 [126] and Mask-RCNN [399] are evaluated on the DIOR dataset. YoloV3 is designed for light-weight and real-time object detection. Mask-RCNN is an extension of Faster-RCNN [124] with ROI align and a third segmentation branch. The experimental results reveal that Mask-RCNN performs better than YoloV3 on this dataset. With regard to different backbones, the results clearly show that Swin-Tiny and ConvNext can outperform other compared methods. The Mask-RCNN method with ConvNext backbone achieves an mAP of 72.4%, which is 6.3% higher than the best results reported in [15]. Notably, we observe that our benchmarked results can greatly outperform the same method reported in existing work. This comparison reveals that the choice of the optimizer, hyper-parameters or other training tricks can greatly affect the final results even when using the same method.

Comparisons on the SEASONET Dataset. SEASONET is a large-scale multi-label LULC scene understanding dataset. It includes 1,759,830 images from Sentinel-2 tiles, and can be used for scene classification, segmentation and retrieval tasks. In this study, we evaluate the segmentation performance on this dataset. On this dataset, we evaluate the widely-used semantic segmentation method DeeplabV3 [392] with three different backbones: ResNet-50, ConvNext, and Swin-Tiny. Upernet and SegFormer [394] with mixed-Transformer encoders (MiT) are also compared. Table 9 reports the benchmarking results. It can be seen that ResNet-50 and Swin-Tiny obtain comparable results and clearly surpass other backbones. SegFormer with MiT encoder clearly outperforms other models. We can also find that the results obtained using EarthNets can significantly outperform performance reported in existing work [21].

Comparisons on the GeoNRW Dataset. The benchmarking results on the GeoNRW dataset are displayed in Table 10. Five segmentation methods, namely, FCN [400], DeeplabV3+ [401], PSPNet [402], Upernet [393] and SegFormer [394] with mixed-Transformer encoders (MiT) are

TABLE 10

Semantic segmentation results on the GeoNRW [286] dataset. We report the aAcc, mAcc and mIoU metrics. The best results are in bold.

Method	Semantic Segmentation				
	Backbone	Optimizer	#Epochs	aAcc	mAcc
MultiTask* [396]	Transfomer	AdamW	100k	76.75	71.89
Lu et al.* [396]	Transfomer	AdamW	100k	76.53	70.12
FCN	UNet	SGD	40k	78.8	66.86
PSPNet	ResNet-50	SGD	40k	81.56	74.92
Deeplabv3+	ResNet-50	SGD	40k	81.91	75.26
Deeplabv3+	ConvNext-Tiny	AdamW	40k	80.89	73.61
Deeplabv3+	ConvNext-Tiny	SGD	40k	78.09	70.01
Deeplabv3+	Swin-Tiny	AdamW	40k	81.11	74.28
Upernet	Swin-Tiny	AdamW	40k	81.87	75.7
Upernet	ResNet-50	SGD	40k	81.87	63.1
Upernet	ConvNext-Tiny	AdamW	40k	82.08	74.9
Upernet	Vit-Small	AdamW	40k	78.65	71.13
Upernet	Swin-Tiny	AdamW	40k	82.31	75.68
SegFormer	MiT	AdamW	40k	82.55	75.63
					64.38

evaluated on this Dataset. we can observe that Transformer-based models like SegFormer and Swin Transformer perform better than other methods. However, the performance of ViT-Small is worse than ResNet-50. When comparing with the reported results in existing work, we can find that using the EarthNets platform can obtain clearly better performance.

7 CONCLUSION

In this study, we present a comprehensive review and build a taxonomy for more than 400 publicly published datasets in the remote sensing community. Based on the attribute information of these datasets, we systemically analyze them from five aspects, including volumes, bibliometric analysis, research domains and the correlation between datasets. Next, a new benchmark including five selected large-scale datasets is built for model evaluation. A deep learning platform termed EarthNets is released towards a consistent evaluation of deep learning methods on remote sensing data. We further use the EarthNets platform to benchmark state-of-the-art methods on the new benchmark. The performance comparisons are insightful for future research.

REFERENCES

- [1] C. Toth and G. Józków, "Remote sensing platforms and sensors: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 22–36, 2016.
- [2] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.
- [3] A. Shaker, W. Y. Yan, and P. E. LaRocque, "Automatic land-water classification using multispectral airborne lidar data for near-shore and river environments," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 94–108, 2019.
- [4] M. E. Bauer, "Remote sensing of environment: history, philosophy, approach and contributions, 1969–2019," *Remote Sensing of Environment*, vol. 237, p. 111522, 2020.
- [5] M. Wójtowicz, A. Wójtowicz, J. Piekarczyk *et al.*, "Application of remote sensing methods in agriculture," *Communications in Biometry and Crop Science*, vol. 11, no. 1, pp. 31–50, 2016.
- [6] L. Karthikeyan, I. Chawla, and A. K. Mishra, "A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses," *Journal of Hydrology*, vol. 586, p. 124905, 2020.
- [7] K. E. Joyce, K. C. Wright, S. V. Samsonov, and V. G. Ambrosia, "Remote sensing and the disaster management cycle," *Advances in geoscience and remote sensing*, vol. 48, p. 7, 2009.
- [8] C. Van Westen, "Remote sensing for natural disaster management," *International archives of photogrammetry and remote sensing*, vol. 33, no. B7/4; PART 7, pp. 1609–1617, 2000.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [10] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [11] Z. Xiong, Y. Yuan, and Q. Wang, "Ai-net: Attention inception neural networks for hyperspectral image classification," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 2647–2650.
- [12] A. Hecheltjen, F. Thonfeld, and G. Menz, "Recent advances in remote sensing change detection—a review," *Land use and land cover mapping in Europe*, pp. 145–178, 2014.
- [13] M. Rahnmemoofar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, vol. 9, pp. 89 644–89 654, 2021.
- [14] Y. Long, G.-S. Xia, S. Li, W. Yang, M. Y. Yang, X. X. Zhu, L. Zhang, and D. Li, "On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-aid," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 14, pp. 4205–4230, 2021.
- [15] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 296–307, 2020.
- [16] A. Abdollahi, B. Pradhan, N. Shukla, S. Chakraborty, and A. Alamri, "Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review," *Remote Sensing*, vol. 12, no. 9, p. 1444, 2020.
- [17] I. Tomljenovic, B. Höfle, D. Tiede, and T. Blaschke, "Building extraction from airborne laser scanning data: An analysis of the state of the art," *Remote Sensing*, vol. 7, no. 4, pp. 3826–3862, 2015.
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [19] X. X. Zhu, J. Hu, C. Qiu, Y. Shi, J. Kang, L. Mou, H. Bagheri, M. Häberle, Y. Hua, R. Huang *et al.*, "So2sat lcz42: A benchmark dataset for global local climate zones classification," *arXiv preprint arXiv:1912.12171*, 2019.
- [20] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, "Functional map of the world," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6172–6180.
- [21] D. Koßmann, V. Brack, and T. Wilhelm, "Seasonet: A seasonal scene classification, segmentation and retrieval dataset for satellite imagery over germany," in *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 243–246.
- [22] A. J. Stewart, C. Robinson, I. A. Corley, A. Ortiz, J. M. L. Ferres, and A. Banerjee, "Torchgeo: deep learning with geospatial data," *arXiv preprint arXiv:2111.08872*, 2021.
- [23] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.
- [24] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.
- [25] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [26] S. Lobry, D. Marcos, J. Murray, and D. Tuia, "Rsvqa: Visual question answering for remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8555–8566, 2020.
- [27] Z. Yuan, L. Mou, Z. Xiong, and X. X. Zhu, "Change detection meets visual question answering," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.
- [28] P. Grönquist, C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler, "Deep learning for post-processing ensemble weather forecasts," 2020.
- [29] W. Tan, N. Qin, L. Ma, Y. Li, J. Du, G. Cai, K. Yang, and J. Li, "Toronto-3d: A large-scale mobile lidar dataset for semantic segmentation of urban roadways," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 202–203.
- [30] G. Can, D. Mantegazza, G. Abbate, S. Chappuis, and A. Giusti, "Swiss3dcities: Aerial photogrammetric 3d pointcloud dataset with semantic labels," Dec. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.4390295>
- [31] ———, "Semantic segmentation on swiss3dcities: a benchmark study on aerial photogrammetric 3d pointcloud dataset," *Pattern Recognition Letters*, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865521001938>
- [32] X. Li, C. Li, Z. Tong, A. Lim, J. Yuan, Y. Wu, J. Tang, and R. Huang, "Campus3d: A photogrammetry point cloud benchmark for hierarchical understanding of outdoor scene," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 238–246.
- [33] S. Workman, R. Souvenir, and N. Jacobs, "Wide-area image geolocation with aerial reference imagery," in *Proceedings of*

- the IEEE International Conference on Computer Vision*, 2015, pp. 3961–3969.
- [34] Z. Zheng, Y. Wei, and Y. Yang, “University-1652: A multi-view multi-source benchmark for drone-based geo-localization,” *ACM Multimedia*, 2020.
- [35] Y. Tian, C. Chen, and M. Shah, “Cross-view image matching for geo-localization in urban environments,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3608–3616.
- [36] E. E. Agency. (2021) Air quality e-reporting (aq e-reporting). [Online]. Available: <https://www.eea.europa.eu/data-and-maps/data/aq-reporting-9>
- [37] V. Sarukkai, A. Jain, B. Uzkent, and S. Ermon, “Dataset From: Cloud Removal from Satellite Images using Spatiotemporal Generator Networks,” 2019. [Online]. Available: <https://doi.org/10.7910/DVN/BSETKZ>
- [38] P. Ebel, A. Meraner, M. Schmitt, and X. X. Zhu, “Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5866–5878, 2020.
- [39] S. Ji, P. Dai, M. Lu, and Y. Zhang, “Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 732–748, 2020.
- [40] D. Cerra, M. Pato, K. Alonso, C. Köhler, M. Schneider, R. de los Reyes, E. Carmona, R. Richter, F. Kurz, P. Reinartz *et al.*, “Dlr-hysu—a benchmark dataset for spectral unmixing,” *Remote Sensing*, vol. 13, no. 13, p. 2559, 2021.
- [41] X. Lu, B. Wang, X. Zheng, and X. Li, “Exploring models and data for remote sensing image caption generation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2183–2195.
- [42] Chen-Yang-Liu. (2022) Remote sensing image change captioning (rsicc). [Online]. Available: <https://github.com/Chen-Yang-Liu/RSICC>
- [43] (2022) Hyperspectral datasets. [Online]. Available: http://www.ahu.eus/ccwintco/index.php/Hyperspectral-Remote_Sensing_Scenes
- [44] X. Sun, Y. Lv, Z. Wang, and K. Fu, “Scan: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution sar images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.
- [45] (2015) District of columbia – classified point cloud lidar. [Online]. Available: <https://github.com/awslabs/open-data-docs/tree/main/docs/dc-lidar-2018>
- [46] ISPRS-Contest. (2022) Isprs 2d semantic labeling contest,” accessed on oct. 9, 2022. [Online]. Available: <https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx>
- [47] ——. (2022) Isprs 2d semantic labeling contest,” accessed on oct. 9, 2022. [Online]. Available: <https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx>
- [48] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 44–51.
- [49] L. Kondmann, A. Toker, M. Rußwurm, A. Camero Unzueta, D. Peressuti, G. Milcinski, N. Longépé, P.-P. Mathieu, T. Davis, G. Marchisio *et al.*, “Denethor: The dynamicearthnet dataset for harmonized, inter-operable, analysis-ready, daily crop monitoring from space,” in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.
- [50] D. Sykas, I. Papoutsis, and D. Zografakis, “Sen4agrinet: A harmonized multi-country, multi-temporal benchmark dataset for agricultural earth observation machine learning applications,” in *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 5830–5833.
- [51] E. L. Amanda A. Boatswain Jacques, Abdoulaye Baniré Diallo, “Towards the creation of a canadian land-use dataset for agricultural land classification,” in *42nd Canadian Symposium on Remote Sensing: Understanding Our World: Remote Sensing for a Sustainable Future*, 2021.
- [52] X. Qi, P. Zhu, Y. Wang, L. Zhang, J. Peng, M. Wu, J. Chen, X. Zhao, N. Zang, and P. T. Mathiopoulos, “Mlrsnet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 337–350, 2020.
- [53] Z. Xiao, Y. Long, D. Li, C. Wei, G. Tang, and J. Liu, “High-resolution remote sensing image retrieval based on cnns from a dimensional perspective,” *Remote Sensing*, vol. 9, no. 7, p. 725, 2017.
- [54] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- [55] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, “Bigearthnet: A large-scale benchmark archive for remote sensing image understanding,” in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5901–5904.
- [56] G. Machado, E. Ferreira, K. Nogueira, H. Oliveira, M. Brito, P. H. T. Gama, and J. A. dos Santos, “Airound and cv-brct: Novel multiview datasets for scene classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 488–503, 2020.
- [57] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [58] R. Bahmanyar, D. Espinoza-Molina, and M. Datcu, “Multisensor earth observation image classification based on a multimodal latent dirichlet allocation model,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 459–463, 2018.
- [59] Y. Zhong, Q. Zhu, and L. Zhang, “Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6207–6222, 2015.
- [60] H. Li, X. Dou, C. Tao, Z. Wu, J. Chen, J. Peng, M. Deng, and L. Zhao, “Rsi-cb: A large-scale remote sensing image classification benchmark using crowdsourced data,” *Sensors*, vol. 20, no. 6, p. 1594, 2020.
- [61] (2018) Ships in satellite imagery. [Online]. Available: <https://www.kaggle.com/datasets/rhammell/ships-in-satellite-imagery>
- [62] A. P. Antonio-Javier Gallego and P. Gil, “Automatic ship classification from optical aerial images with convolutional neural networks,” *Remote Sensing*, vol. 10, no. 4, 2018.
- [63] Y. Di, Z. Jiang, H. Zhang, and G. Meng, “A public dataset for ship classification in remote sensing images,” in *Image and Signal Processing for Remote Sensing XXV*, vol. 11155. SPIE, 2019, pp. 515–521.
- [64] I. G. Rizaev and A. Achim, “Synthwakesar: A synthetic sar dataset for deep learning classification of ships at sea,” *Remote Sensing*, vol. 14, no. 16, p. 3999, 2022.
- [65] Q. D. Cao and Y. Choe, “Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks,” *Natural Hazards*, vol. 103, no. 3, pp. 3357–3376, 2020.
- [66] C. Rambour, N. Audebert, E. Koeniguer, B. Le Saux, M. Crucianu, and M. Datcu, “Flood detection in time series of optical and sar images,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 1343–1346, 2020.
- [67] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, “Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 210–211.
- [68] R. Ba, C. Chen, J. Yuan, W. Song, and S. Lo, “Smokenet: Satellite smoke scene detection using convolutional neural network with spatial and channel-wise attention,” *Remote Sensing*, vol. 11, no. 14, p. 1702, 2019.
- [69] (2017) Noaa fisheries steller sea lion population count. [Online]. Available: <https://www.kaggle.com/c/noaa-fisheries-steller-sea-lion-population-count/data>
- [70] N. Imamoglu, M. Kimura, H. Miyamoto, A. Fujita, and R. Nakamura, “Solar power plant detection on multi-spectral satellite imagery using weakly-supervised cnn with feedback features and m-pcnn fusion,” *arXiv preprint arXiv:1704.06410*, 2017.
- [71] (2018) Airbus wind turbines patches. [Online]. Available: <https://www.kaggle.com/datasets/airbusgeo/airbus-wind-turbines-patches>
- [72] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, “220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3,” *Purdue University Research Repository*, vol. 10, no. 7, p. 991, 2015.

- [73] R. M Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, "Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 75–82.
- [74] H. Kerner, C. Nakalembe, and I. Becker-Reshef, "Field-level crop type classification with k nearest neighbors: a baseline for a new kenya smallholder dataset," *arXiv preprint arXiv:2004.03023*, 2020.
- [75] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefèvre, and M. Körner, "Breizhcrobs: A time series dataset for crop type mapping," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS* (2020), 2020.
- [76] S. Virnodkar, V. Pachghare, V. Patil, and S. K. Jha, "Canesat," 2020. [Online]. Available: <https://dx.doi.org/10.21227/vzbn-qj64>
- [77] G. Tseng, I. Zvonkov, C. L. Nakalembe, and H. Kerner, "Cropharvest: A global dataset for crop-type classification," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.
- [78] R. E. F. Western Cape Department of Agricultures. (2021) Crop type classification dataset for western cape, south africa. [Online]. Available: <https://doi.org/10.34911/rdnt.j0co8q>
- [79] G. Choumos, A. Koukos, V. Sitokonstantinou, and C. Kontoes, "Towards space-to-ground data availability for agriculture monitoring," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2022, pp. 1–5.
- [80] M. J. Hughes and D. J. Hayes, "Automated detection of cloud and cloud shadow in single-date landsat imagery using neural networks and spatial post-processing," *Remote Sensing*, vol. 6, no. 6, pp. 4907–4926, 2014.
- [81] M. P. I. for Meteorology. (2019) Understanding clouds from satellite images. [Online]. Available: https://www.kaggle.com/c/understanding_cloud_organization/data
- [82] A. H. Nielsen, A. Iosifidis, and H. Karstoft, "Cloudcast: A satellite-based dataset and baseline for forecasting clouds," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3485–3494, 2021.
- [83] (2020) Sentinel-2 cloud mask catalogue. [Online]. Available: <https://zenodo.org/record/4172871#.Y0LEOHbP1D->
- [84] L. Mou, Y. Hua, P. Jin, and X. X. Zhu, "ERA: A dataset and deep learning benchmark for event recognition in aerial videos," *IEEE Geoscience and Remote Sensing Magazine*, in press.
- [85] G. I. Drakonakis, G. Tsagkatakis, K. Fotiadou, and P. Tsakalides, "Ombrianner—supervised flood mapping via convolutional neural networks using multitemporal sentinel-1 and sentinel-2 data fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2341–2356, 2022.
- [86] S. Chandak, V. Chittlers, and S. Honnunyar, "Understanding the amazon rainforest from space using cnns," 2017.
- [87] D. Noever and S. E. M. Noever, "Overhead mnist: A benchmark satellite dataset," *arXiv preprint arXiv:2102.04266*, 2021.
- [88] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maître, "Structural high-resolution satellite image indexing," Vienna, Austria, 2010.
- [89] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 2015.
- [90] L. Zhao, P. Tang, and L. Huo, "Feature significance-based multibag-of-visual-words model for remote sensing image scene classification," *Journal of Applied Remote Sensing*, vol. 10, no. 3, p. 035004, 2016.
- [91] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 197–209, 2018.
- [92] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of vhr remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1155–1167, 2018.
- [93] H. Li, H. Jiang, X. Gu, J. Peng, W. Li, L. Hong, and C. Tao, "Clrs: Continual learning benchmark for remote sensing image scene classification," *Sensors*, vol. 20, no. 4, p. 1226, 2020.
- [94] Z. Zhou, S. Li, W. Wu, W. Guo, X. Li, G. Xia, and Z. Zhao, "Nasc-tg2: Natural scene classification with tiangong-2 remotely sensed imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3228–3242, 2021.
- [95] M. Reda. (2021) Satellite image classification. [Online]. Available: <https://www.kaggle.com/datasets/mahmoudreda55/satellite-image-classification>
- [96] Y. Hua, L. Mou, P. Jin, and X. X. Zhu, "Multiscene: A large-scale dataset and benchmark for multiscene recognition in single aerial images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.
- [97] N. I. Bountos, I. Papoutsis, D. Michail, A. Karavias, P. Elias, and I. Parcharidis, "Hephaestus: A large scale multitask dataset towards insar understanding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1453–1462.
- [98] K. Uehara, H. Sakanashi, H. Nosato, M. Murakawa, H. Miyamoto, and R. Nakamura, "Object detection of satellite images using multi-channel higher-order local autocorrelation," in *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, 2017, pp. 1339–1344.
- [99] Statoil. (2017) Statoil/c-core iceberg classifier challenge. [Online]. Available: <https://www.kaggle.com/competitions/statoil-iceberg-classifier-challenge>
- [100] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat: a learning framework for satellite imagery," in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 2015, pp. 1–10.
- [101] (2017) Tiselac: Time series land cover classification challeng. [Online]. Available: <https://sites.google.com/site/dinoienco/tiselc>
- [102] X.-Y. Tong, G.-S. Xia, Q. Lu, H. Shen, S. Li, S. You, and L. Zhang, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sensing of Environment*, vol. 237, p. 111322, 2020.
- [103] (2019) Eopatches for slovenia 2019. [Online]. Available: <http://eo-learn.sentinel-hub.com/>
- [104] (2016) Tiangong1 high-resolution satellite scene classification. [Online]. Available: <http://www.msadc.cn/main/setsubDetail?id=1369487569196158978>
- [105] (2017) Austin zoning satellite images. [Online]. Available: <https://www.kaggle.com/datasets/franchenstein/austin-zoning-satellite-images>
- [106] R. Ratajczak, C. F. Crispim-Junior, É. Faure, B. Fervers, and L. Tougne, "Automatic Land Cover Reconstruction From Historical Aerial Images: An Evaluation of Features Extraction and Classification Algorithms," *IEEE Transactions on Image Processing*, Jan. 2019.
- [107] (2020) Sense earth satellite scene classification. [Online]. Available: <https://aistudio.baidu.com/aistudio/datasetdetail/52728>
- [108] S. Ji, D. Yu, C. Shen, W. Li, and Q. Xu, "Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks," *Landslides*, vol. 17, no. 6, pp. 1337–1352, 2020.
- [109] U. S. Ranjan and A. Narayana, "Classification of objects in sar images using scaling features," in *ICVGIP*, 2002.
- [110] (2018) Forest cover type (kernels only). [Online]. Available: <https://www.kaggle.com/c/forest-cover-type-kernels-only/data>
- [111] (2019) Aerial cactus identification. [Online]. Available: <https://www.kaggle.com/c/aerial-cactus-identification>
- [112] (2019) Wids datathon 2019. [Online]. Available: <https://www.kaggle.com/c/widsdatathon2019>
- [113] S. Beery, G. Wu, T. Edwards, F. Pavetic, B. Majewski, S. Mukherjee, S. Chan, J. Morgan, V. Rathod, and J. Huang, "The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 294–21 307.
- [114] S. Ahlsweide, C. Schulz, C. Gava, P. Helber, B. Bischke, M. Förster, F. Arias, J. Hees, B. Demir, and B. Kleinschmit, "Treesatai benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing," *Earth System Science Data Discussions*, pp. 1–22, 2022.
- [115] S. F. Agency. (2021) Forest damages – larch casebearer 1.0. national forest data lab. dataset. [Online]. Available: <https://lila.science/datasets/forest-damages-larch-casebearer>
- [116] E. Cole, B. Deneu, T. Loriel, M. Servajean, C. Botella, D. Morris, N. Jovic, P. Bonnet, and A. Joly, "The geolifecl 2020 dataset," *arXiv preprint arXiv:2004.04192*, 2020.

- [117] E. Ferreira, M. Brito, R. Balaniuk, M. S. Alvim, and J. A. dos Santos, "Brazildam: A benchmark dataset for tailings dam detection," in *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*. IEEE, 2020, pp. 339–344.
- [118] B. Chen, Q. Feng, B. Niu, F. Yan, B. Gao, J. Yang, J. Gong, and J. Liu, "Multi-modal fusion of satellite and street-view images for urban village classification based on a dual-branch deep neural network," *International Journal of Applied Earth Observation and Geoinformation*, vol. 109, p. 102794, 2022.
- [119] (2022) Kennedy space center. [Online]. Available: <https://www.csr.utexas.edu/projects/rs/hrs/classify.html>
- [120] K. Nogueira, J. A. Dos Santos, T. Fornazari, T. S. F. Silva, L. P. Morellato, and R. d. S. Torres, "Towards vegetation species discrimination by using data-driven descriptors," in *2016 9th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS)*. Ieee, 2016, pp. 1–6.
- [121] B. Uzkent, A. Rangnekar, and M. J. Hoffman, "Tracking in aerial hyperspectral videos using deep kernelized correlation filters," *arXiv preprint arXiv:1711.07235*, 2017.
- [122] (2019) Find a car park. [Online]. Available: <https://www.kaggle.com/datasets/daggysheep/find-a-car-park>
- [123] E. Dahan, T. Diskin, A. Amram, A. Moryossef, and O. Koren, "Cofga: A dataset for fine grained classification of objects from aerial imagery," *arXiv preprint arXiv:2105.12786*, 2021.
- [124] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [125] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.
- [126] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [127] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.
- [128] RHAMMELL. (2017) Planes in satellite imagery. [Online]. Available: <https://www.kaggle.com/datasets/rhammell/planesnet>
- [129] B. A. Jeff Faudi. (2021) Airbus aircraft detection. [Online]. Available: <https://www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset>
- [130] S. Rawat. (2021) Casia-aircraft. [Online]. Available: <https://www.rsaicp.com/portal/dataDetail?id=16>
- [131] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.
- [132] Microsoft. (2019) Us building footprints. [Online]. Available: <https://github.com/microsoft/USBuildingFootprints>
- [133] (2018) Segmenting buildings in satellite images. [Online]. Available: <https://www.kaggle.com/code/kmader/segmenting-buildings-in-satellite-images/data>
- [134] N. Weir, D. Lindenbaum, A. Bastidas, A. V. Etten, S. McPherson, J. Shermeyer, V. Kumar, and H. Tang, "Spacenet mvoi: A multi-view overhead imagery dataset," in *Proceedings of the ieee/cvf international conference on computer vision*, 2019, pp. 992–1001.
- [135] L. Huang, B. Liu, B. Li, W. Guo, W. Yu, Z. Zhang, and W. Yu, "Opensarship: A dataset dedicated to sentinel-1 ship interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 195–208, 2017.
- [136] J. Li, C. Qu, and J. Shao, "Ship detection in sar images based on an improved faster r-cnn," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*. IEEE, 2017, pp. 1–6.
- [137] Airbus. (2018) Airbus ship detection challenge. [Online]. Available: <https://www.kaggle.com/c/airbus-ship-detection>
- [138] p. m. adrian and A. Umam. (2018) Ships in google earth. [Online]. Available: <https://www.kaggle.com/datasets/tomluther/ships-in-google-earth>
- [139] G. Heitz and D. Koller, "Learning spatial context: Using stuff to find things," in *European conference on computer vision*. Springer, 2008, pp. 30–43.
- [140] S. Razkarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016.
- [141] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A large contextual dataset for classification, detection and counting of cars with deep learning," in *European conference on computer vision*. Springer, 2016, pp. 785–800.
- [142] DLR. (2016) Dlr multi-class vehicle detection and orientation in aerial imagery. [Online]. Available: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12760/22294_read-52777
- [143] M. Y. Yang, W. Liao, X. Li, and B. Rosenhahn, "Deep learning for vehicle detection in aerial images," in *IEEE International Conference on Image Processing (ICIP)*, 2018.
- [144] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote sensing*, vol. 117, pp. 11–28, 2016.
- [145] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2486–2498, 2017.
- [146] Y. Zhang, Y. Yuan, Y. Feng, and X. Lu, "Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5535–5548, 2019.
- [147] xView. (2019) xview 2018 object detection challenge. [Online]. Available: <https://challenge.xviewdataset.org/welcome>
- [148] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, "Functional map of the world," in *CVPR*, 2018.
- [149] W. Han, J. Li, S. Wang, Y. Wang, J. Yan, R. Fan, X. Zhang, and L. Wang, "A context-scale-aware detector and a new benchmark for remote sensing small weak object detection in unmanned aerial vehicle images," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 112, p. 102966, 2022.
- [150] (2017) Dstl satellite imagery feature detection. [Online]. Available: https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection/data?select=three_band.zip
- [151] Airbus. (2021) Airbus oil storage detection. [Online]. Available: <https://www.kaggle.com/datasets/airbusgeo/airbus-oil-storage-detection-dataset>
- [152] T. Zhang. (2020) Large-scale sar ship detection dataset-v1.0. [Online]. Available: <https://github.com/TianwenZhang0825/LS-SSDD-v1.0-OPEN>
- [153] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "Hrsid: A high-resolution sar images dataset for ship detection and instance segmentation," *Iee Access*, vol. 8, pp. 120234–120254, 2020.
- [154] J. Wang, W. Yang, H. Guo, R. Zhang, and G.-S. Xia, "Tiny object detection in aerial images," 2021, pp. 3791–3798.
- [155] G. Cheng, X. Yuan, X. Yao, K. Yan, Q. Zeng, and J. Han, "Towards large-scale small object detection: Survey and benchmarks," *arXiv preprint arXiv:2207.14096*, 2022.
- [156] (2022) Hydice image of washington dc mall. [Online]. Available: <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>
- [157] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271618301229>
- [158] A. Boguszewski, D. Batorski, N. Ziembka-Jankowska, T. Dziedzic, and A. Zambrzycka, "Landcover. ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1102–1110.
- [159] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms—a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *arXiv preprint arXiv:1906.07789*, 2019.
- [160] (2019) Semantic segmentation of aerial imagery. [Online]. Available: <https://www.kaggle.com/datasets/humansintheloop/semantic-segmentation-of-aerial-imagery>
- [161] Z. Wang, X. Zeng, Z. Yan, J. Kang, and X. Sun, "Air-polsar-seg: A large-scale data set for terrain segmentation in complex-scene polsar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3830–3841, 2022.
- [162] V. Sainte Fare Garnot and L. Landrieu, "Panoptic segmentation of satellite image time series with convolutional temporal attention networks," *ICCV*, 2021.

- [163] Z. Wu, "Multi-type aircraft of remote sensing images: Mtarsi," 2019. [Online]. Available: <https://github.com/azavea/cloud-model>
- [164] (2020) Rareplanes. [Online]. Available: <https://www.cosmiquworks.org/rareplanes-public-user-guide/>
- [165] S. Rawat. (2021) Cgi planes in satellite imagery w/ bboxes. [Online]. Available: <https://www.kaggle.com/datasets/aceofspades914/cgi-planes-in-satellite-imagery-w-bboxes>
- [166] P. Zhang, H. Xu, T. Tian, P. Gao, L. Li, T. Zhao, N. Zhang, and J. Tian, "Sefepnet: Scale expansion and feature enhancement pyramid network for sar aircraft detection with small sample dataset," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3365–3375, 2022.
- [167] W. Yu, G. Cheng, M. Wang, Y. Yao, X. Xie, X. Yao, and J. Han, "Mar20: A benchmark for military aircraft recognition in remote sensing images," *National Remote Sensing Bulletin*.
- [168] K. Nogueira, C. da Silva, P. Gama, G. Machado, and J. A. Dos Santos, "A tool for bridge detection in major infrastructure works using satellite images." *2019 Workshop of Computer Vision (WVC)*, 2019.
- [169] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 574–586, 2018.
- [170] S. P. Mohanty, J. Czakon, K. A. Kaczmarek, A. Pyskir, P. Tarasiewicz, S. Kunwar, J. Rohrbach, D. Luo, M. Prasad, S. Fleer et al., "Deep learning for understanding satellite imagery: An experimental survey," *Frontiers in Artificial Intelligence*, vol. 3, 2020.
- [171] (2018) 2018 open ai tanzania building footprint segmentation challenge. [Online]. Available: <https://competitions.codalab.org/competitions/20100>
- [172] Microsoft. (2019) Australia building footprints. [Online]. Available: <https://github.com/microsoft/AustraliaBuildingFootprints>
- [173] ——. (2019) Uganda tanzania building footprints public. [Online]. Available: <https://github.com/microsoft/Uganda-Tanzania-Building-Footprints>
- [174] ——. (2019) Canadian building footprints. [Online]. Available: <https://github.com/Microsoft/CanadianBuildingFootprints>
- [175] AICyberTeam. (2022) Urban building classification dataset. [Online]. Available: <https://github.com/AICyberTeam/UBC-dataset>
- [176] J. Wang, L. Meng, W. Li, W. Yang, L. Yu, and G.-S. Xia, "Learning to extract building footprints from off-nadir aerial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2022.
- [177] Y. L. G.-S. X. Q. L. Jian Ding, Nan Xue, "Learning roi transformer for detecting oriented objects in aerial images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [178] J. Ding, N. Xue, G.-S. Xia, X. Bai, W. Yang, M. Yang, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.
- [179] S. Waqas Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. Shahbaz Khan, F. Zhu, L. Shao, G.-S. Xia, and X. Bai, "isaид: A large-scale dataset for instance segmentation in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 28–37.
- [180] VALID. (2020) Virtual aerial image dataset. [Online]. Available: <https://sites.google.com/view/valid-dataset>
- [181] E. Bondi, R. Jain, P. Aggrawal, S. Anand, R. Hannaford, A. Kapoor, J. Piavis, S. Shah, L. Joppa, B. Dilkina et al., "Birdsai: A dataset for detection and tracking in aerial thermal infrared videos," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1747–1756.
- [182] G. Gao, Q. Liu, and Y. Wang, "Counting from sky: A large-scale data set for remote sensing object counting and a benchmark method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3642–3655, 2020.
- [183] S. C. D. Rabbi, Jakaria; Chowdhury, "Oil and gas tank dataset," 2020. [Online]. Available: <https://data.mendeley.com/datasets/bkxj8z84m9/3>
- [184] H. Karl and M. Md. (2019) Oil storage tanks. [Online]. Available: <https://www.kaggle.com/datasets/towardsentropy/oil-storage-tanks>
- [185] N. I. Bountos, I. Papoutsis, D. Michail, A. Karavias, P. Elias, and I. Parcharidis, "Hephaestus: A large scale multitask dataset towards insar understanding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2022, pp. 1453–1462.
- [186] (2019) Semantic drone dataset. [Online]. Available: <https://www.tugraz.at/index.php?id=22387>
- [187] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory prediction in crowded scenes," in *European Conference on Computer Vision (ECCV)*, vol. 2, 2020.
- [188] J. Solawetz. (2020) Aerial maritime drone dataset. [Online]. Available: <https://public.roboflow.com/object-detection/aerial-maritime>
- [189] L. A. Varga, B. Kiefer, M. Messmer, and A. Zell, "Seadronesee: A maritime benchmark for detecting humans in open water," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2260–2270.
- [190] J. Gasienica-Jozkowy, M. Knapik, and B. Cyganek, "An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance," *Integrated Computer-Aided Engineering*, pp. 1–15, 01 2021.
- [191] M. Thoreau and F. Wilson, "Sarnet: A dataset for deep learning assisted search and rescue with satellite imagery," 2021.
- [192] Z. Liu, L. Yuan, L. Weng, and Y. Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *International conference on pattern recognition applications and methods*, vol. 2. SciTePress, 2017, pp. 324–331.
- [193] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A sar dataset of ship detection for deep learning under complex backgrounds," *Remote Sensing*, vol. 11, no. 7, 2019. [Online]. Available: <http://www.mdpi.com/2072-4292/11/7/765>
- [194] S. Xian, W. Zhirui, S. Yuanrui, D. Wenhui, Z. Yue, and F. Kun, "Air-sarship-1.0: High-resolution sar ship detection dataset," *Journal of Radars*, vol. 8, no. 6, pp. 852–863, 2019.
- [195] F. Xue, W. Jin, S. Qiu, and J. Yang, "Rethinking automatic ship wake detection: State-of-the-art cnn-based wake detection via optical images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–22, 2021.
- [196] (2020) Casia-ship. [Online]. Available: <https://www.rsaicp.com/portal/dataDetail?id=14>
- [197] xvview. (2022) xvview3 competition. [Online]. Available: <https://iuiu.xvview.us/>
- [198] I. Bozcan and E. Kayacan, "Au-air: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8504–8510.
- [199] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The hignd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2118–2125.
- [200] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kümmeler, H. Königshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "INTERACTION Dataset: An INTERnational, Adversarial and Cooperative motion Dataset in Interactive Driving Scenarios with Semantic Maps," *arXiv:1910.03088 [cs, eess]*, 2019.
- [201] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 1929–1934.
- [202] R. Krajewski, T. Moers, J. Bock, L. Vater, and L. Eckstein, "The round dataset: A drone dataset of road user trajectories at roundabouts in germany," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.
- [203] B. Weinstein, S. Marconi, A. Zare, S. Bohlman, S. Graves, A. Singh, and E. White, "Neon tree crowns dataset," Apr. 2020, gordon and Betty Moore Foundation: GBMF4563. [Online]. Available: <https://doi.org/10.5281/zenodo.3765872>
- [204] S. F. Agency. (2021) Forest damages – larch casebearer 1.0. [Online]. Available: <https://lila.science/datasets/forest-damages-larch-casebearer/>
- [205] (2009) Overhead imagery research data set. [Online]. Available: <https://sourceforge.net/projects/oirds/>

- [206] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao, "Orientation robust object detection in aerial images using deep convolutional neural network," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 3735–3739.
- [207] P. R. De Almeida, L. S. Oliveira, A. S. Britto Jr, E. J. Silva Jr, and A. L. Koerich, "Pklot—a robust dataset for parking lot classification," *Expert Systems with Applications*, vol. 42, no. 11, pp. 4937–4949, 2015.
- [208] M.-R. Hsieh, Y.-L. Lin, and W. H. Hsu, "Drone-based object counting by spatially regularized regional proposal network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4145–4153.
- [209] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.
- [210] M. Haroon, M. Shahzad, and M. M. Fraz, "Multi-sized object detection using spaceborne optical imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3032–3046, 2020.
- [211] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [212] S. M. Azimi, R. Bahmanyar, C. Henry, and F. Kurz, "Eagle: Large-scale vehicle detection dataset in real-world scenarios using aerial imagery," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 6920–6927.
- [213] M. Mandal, L. K. Kumar, and S. K. Vipparthi, "Mor-uav: A benchmark dataset and baselines for moving object recognition in uav videos," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2626–2635.
- [214] Y. Sun, B. Cao, P. Zhu, and Q. Hu, "Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2022.
- [215] Ari, M. Enxhi, and K. Elyes. (2019) Swimming pool and car detection. [Online]. Available: <https://www.kaggle.com/datasets/kbhartiya83/swimming-pool-and-car-detection>
- [216] I. Weber, J. Bongartz, and R. Roscher, "Artifive-potsdam: A benchmark for learning with artificial objects for improved aerial vehicle detection," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 1214–1217.
- [217] Y. Liu, J. Yao, X. Lu, M. Xia, X. Wang, and Y. Liu, "Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2043–2056, 2019.
- [218] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 3226–3229.
- [219] G. Mátyus, S. Wang, S. Fidler, and R. Urtasun, "Hd maps: Fine-grained road segmentation by parsing ground and aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3611–3619.
- [220] G. Mátyus, S. Wang, S. Fidler, and R. Urtasun, "Enhancing road maps by parsing aerial images around the world," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1689–1697.
- [221] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden, and D. DeWitt, "Roadtracer: Automatic extraction of road networks from aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4720–4728.
- [222] (2020) Microsoft roaddetections. [Online]. Available: <https://github.com/microsoft/RoadDetections>
- [223] S. Foga, P. L. Scaramuzza, S. Guo, Z. Zhu, R. D. Dilley Jr, T. Beckmann, G. L. Schmidt, J. L. Dwyer, M. J. Hughes, and B. Laue, "Cloud detection algorithm comparison and validation for operational landsat data products," *Remote sensing of environment*, vol. 194, pp. 379–390, 2017.
- [224] S. Mohajerani and P. Saeedi, "Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, July 2019, pp. 1029–1032.
- [225] S. Mohajerani and P. Saeedi, "Cloud-Net+: A Cloud Segmentation CNN for Landsat 8 Remote Sensing Imagery Optimized with Filtered Jaccard Loss Function," vol. 2001.08768, 2020.
- [226] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for remote sensing images by the fusion of multi-scale convolutional features," *arXiv preprint arXiv:1810.05801*, 2018.
- [227] J. Li, Z. Wu, Z. Hu, C. Jian, S. Luo, L. Mou, X. X. Zhu, and M. Molinier, "A lightweight deep learning-based cloud detection method for sentinel-2a imagery fusing multiscale spectral and spatial features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2021.
- [228] J. A. Barsi, K. Lee, G. Kváran, B. L. Markham, and J. A. Pedelty, "The spectral response of the landsat-8 operational land imager," *Remote sensing*, vol. 6, no. 10, pp. 10 232–10 251, 2014.
- [229] S. Ji, Z. Zhang, C. Zhang, S. Wei, M. Lu, and Y. Duan, "Learning discriminative spatiotemporal features for precise crop classification from multi-temporal satellite images," *International Journal of Remote Sensing*, vol. 41, no. 8, pp. 3162–3174, 2020.
- [230] G. Weikmann, C. Paris, and L. Bruzzone, "Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4699–4708, 2021.
- [231] M. T. Chiu, X. Xu, Y. Wei, Z. Huang, A. G. Schwing, R. Brunner, H. Khachatrian, H. Karapetyan, I. Dozier, G. Rose et al., "Agriculture-vision: A large aerial image database for agricultural pattern analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2828–2838.
- [232] I. de Gélis, S. Lefèvre, and T. Corpetti, "Change detection in urban point clouds: An experimental comparison with simulated 3d datasets," *Remote Sensing*, vol. 13, no. 13, p. 2629, 2021.
- [233] L. Shen, Y. Lu, H. Chen, H. Wei, D. Xie, J. Yue, R. Chen, S. Lv, and B. Jiang, "S2looking: A satellite side-looking dataset for building change detection," *Remote Sensing*, vol. 13, no. 24, p. 5094, 2021.
- [234] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *arXiv preprint arXiv:1909.05948*, 2019.
- [235] K. Yang, G.-S. Xia, Z. Liu, B. Du, W. Yang, M. Pelillo, and L. Zhang, "Asymmetric siamese networks for semantic change detection in aerial images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2021.
- [236] C. F. Brown, S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko et al., "Dynamic world, near real-time global 10 m land use land cover mapping," *Scientific Data*, vol. 9, no. 1, pp. 1–17, 2022.
- [237] T. P. Sonali Patil, Bharath Comandur and A. C. Kak, "A New Stereo Benchmarking Dataset for Satellite Images," *arXiv:1907.04404*, 2019. [Online]. Available: <http://arxiv.org/abs/1907.04404>
- [238] J. Liu and S. Ji, "A novel recurrent encoder-decoder structure for large-scale multi-view stereo reconstruction from an open aerial dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6050–6059.
- [239] M. Kölle, D. Laupheimer, S. Schmohl, N. Haala, F. Rottensteiner, J. D. Wegner, and H. Ledoux, "The hessigheim 3d (h3d) benchmark on semantic segmentation of high-resolution 3d point clouds and textured meshes from uav lidar and multi-view-stereo," *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 1, p. 100001, 2021.
- [240] K. Kashinath, M. Mudigonda, S. Kim, L. Kapp-Schwoerer, A. Graubner, E. Karaismailoglu, L. Von Kleist, T. Kurth, A. Greiner, A. Mahesh et al., "Climatenet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather," *Geoscientific Model Development*, vol. 14, no. 1, pp. 107–124, 2021.
- [241] (2022) Hyperview challenge. [Online]. Available: <https://platform.ai4eo.eu/seeing-beyond-the-visible>
- [242] R. E. Foundation. (2022) Tropical cyclone wind estimation competition dataset, version 1.0, radiant mlhub. [Online]. Available: <https://doi.org/10.34911/rdnt.xs53up>
- [243] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CVPR*, 2017.
- [244] O. Mañas, A. Lacoste, X. Giro-i Nieto, D. Vazquez, and P. Rodriguez, "Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data," *arXiv preprint arXiv:2103.16607*, 2021.

- [245] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "Whuhi: Uav-borne hyperspectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf," *Remote Sensing of Environment*, vol. 250, p. 112012, 2020.
- [246] M. O. Turkoglu, S. D'Aronco, G. Perich, F. Liebisch, C. Streit, K. Schindler, and J. D. Wegner, "Crop mapping from image time series: Deep learning with multi-scale label hierarchies," *Remote Sensing of Environment*, vol. 264, p. 112603, 2021.
- [247] M. Schneider, A. Broszeit, and M. Körner, "EuroCrops: A pan-european dataset for time series crop type classification," in *Proceedings of the Conference on Big Data from Space (BiDS)*, P. Soille, S. Loekken, and S. Albani, Eds. Publications Office of the European Union.
- [248] Y.-R. Wang and X.-M. Li, "Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning," *Earth System Science Data*, vol. 13, no. 6, pp. 2723–2742, 2021.
- [249] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018.
- [250] A. Shakeel, W. Sultani, and M. Ali, "Deep built-structure counting in satellite imagery using attention based re-weighting," *ISPRS journal of photogrammetry and remote sensing*, vol. 151, pp. 313–321, 2019.
- [251] J. Shermeyer, D. Hogan, J. Brown, A. Van Etten, N. Weir, F. Pacifici, R. Hansch, A. Bastidas, S. Soenen, T. Bacastow et al., "Spacenet 6: Multi-sensor all weather mapping dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 196–197.
- [252] A. Van Etten, D. Hogan, J. M. Manso, J. Shermeyer, N. Weir, and R. Lewis, "The multi-temporal urban development spacenet dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6398–6407.
- [253] F. Kong, B. Huang, K. Bradbury, and J. Malof, "The synthinel-1 dataset: A collection of high resolution synthetic overhead imagery for building segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1814–1823.
- [254] V. Mnih, "Machine learning for aerial image labeling," Ph.D. dissertation, University of Toronto, 2013.
- [255] G. Labs. (2018) Open cities ai challenge dataset, version 1.0, radiant mlhub. [Online]. Available: <https://doi.org/10.34911/rdnt.f94cxb>
- [256] G. repository. (2018) Segmenting buildings in satellite images. [Online]. Available: <https://github.com/QiaoWenfan/Building-Dataset>
- [257] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6054–6068, 2017.
- [258] L. Baetens, C. Desjardins, and O. Hagolle, "Validation of copernicus sentinel-2 cloud masks obtained from maja, sen2cor, and fmask processors using reference cloud masks generated with a supervised active learning procedure," *Remote Sensing*, vol. 11, no. 4, p. 433, 2019.
- [259] Q. He, X. Sun, Z. Yan, and K. Fu, "Dabnet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2021.
- [260] G. repository. (2018) The azavea cloud dataset. [Online]. Available: <https://github.com/azavea/cloud-model>
- [261] R. E. Foundation. (2022) Sentinel-2 cloud cover segmentation dataset (version 1). radiant mlhub. [Online]. Available: <https://doi.org/10.34911/rdnt.hfq6m7>
- [262] Z. Shao, K. Yang, and W. Zhou, "Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset," *Remote Sensing*, vol. 10, no. 6, p. 964, 2018.
- [263] J. Yuan, S. S. Gleason, and A. M. Cheriyyadat, "Systematic benchmarking of aerial image segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1527–1531, 2013.
- [264] G. Moser, D. Tuia, and M. Shimoni, "2015 ieee grss data fusion contest: Extremely high resolution lidar and optical data [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 1, pp. 40–41, 2015.
- [265] M. Volpi and V. Ferrari, "Semantic segmentation of urban scenes by learning local class interactions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 1–9.
- [266] (2016) Dstl satellite imagery feature detection. [Online]. Available: <https://www.kaggle.com/competitions/dstl-satellite-imagery-feature-detection/data>
- [267] M. Zhang, X. Hu, L. Zhao, Y. Lv, M. Luo, and S. Pang, "Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images," *Remote Sensing*, vol. 9, no. 5, p. 500, 2017.
- [268] M. Scanlon, "Semantic annotation of aerial images using deep learning, transfer learning, and synthetic training data," Ph.D. dissertation, 09 2018.
- [269] I. Nigam, C. Huang, and D. Ramanan, "Ensemble knowledge transfer for semantic segmentation," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 1499–1508.
- [270] Y. Chen, Y. Wang, P. Lu, Y. Chen, and G. Wang, "Large-scale structure from motion with semantic constraints of aerial images," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. Springer, 2018, pp. 347–359.
- [271] (2019) Drone deploy medium dataset. [Online]. Available: <https://github.com/dronedeploy/dd-ml-segmentation-benchmark>
- [272] M. Fonder and M. Van Droogenbroeck, "Mid-air: A multi-modal dataset for extremely low altitude drone flights," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.
- [273] A. Rangnekar, N. Mokashi, E. J. Lentilucci, C. Kanan, and M. J. Hoffman, "Aerorit: A new scene for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 8116–8124, 2020.
- [274] R. Roscher, M. Volpi, C. Mallet, L. Drees, and J. D. Wegner, "Sem-city toulouse: A benchmark for building instance segmentation in satellite images," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 5, 2020, pp. 109–116.
- [275] Y. Lyu, G. Vosselman, G.-S. Xia, A. Yilmaz, and M. Y. Yang, "Uavid: A semantic segmentation dataset for uav imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 165, pp. 108 – 119, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271620301295>
- [276] (2021) 2021 ieee grss data fusion contest track dse. [Online]. Available: www.grss-ieee.org/community/technical-committees/data-fusion
- [277] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 172–181.
- [278] C. K. Z. A.-G. Karantzalos, K.; Karakizi, "Hyrank hyperspectral satellite dataset i (version v001)," in *I.W.III/4; ISPRS: Hannover, Germany*, 2018.
- [279] B. Le Saux, N. Yokoya, R. Hänsch, and M. Brown, "2019 ieee grss data fusion contest: large-scale semantic 3d reconstruction," *IEEE Geoscience and Remote Sensing Magazine (GRSM)*, vol. 7, no. 4, pp. 33–36, 2019.
- [280] (2019) Aerial hyperspectral remote sensing image dataset of horseshoe bay village in xiongan. [Online]. Available: <http://www.hrs-cas.com/a/share/shujuchanpin/2019/0501/1049.html>
- [281] C. Robinson, L. Hou, K. Malkin, R. Soobitsky, J. Czawltyko, B. Dilkina, and N. Jojic, "Large scale high-resolution land cover mapping with multi-resolution data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12726–12735.
- [282] N. Yokoya, P. Ghamisi, R. Hänsch, and M. Schmitt, "2020 ieee grss data fusion contest: Global land cover mapping with weak supervision [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 1, pp. 154–157, 2020.
- [283] (2019) Ccf big data & computing intelligence contest 2020. [Online]. Available: <https://www.datafountain.cn/competitions/466>
- [284] J. Wang, Z. Zheng, A. Ma, X. Lu, and Y. Zhong, "Loveda: A remote sensing land-cover dataset for domain adaptive semantic segmentation," *arXiv preprint arXiv:2110.08733*, 2021.
- [285] J. Castillo-Navarro, B. Le Saux, A. Boulch, N. Audebert, and S. Lefèvre, "Semi-supervised semantic segmentation in earth observation: The minifrance suite, dataset analysis and multi-task network study," *Machine Learning*, pp. 1–36, 2021.

- [286] G. Baier, A. Deschamps, M. Schmitt, and N. Yokoya, "Synthesizing optical and sar imagery from land cover maps and auxiliary raster data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.
- [287] R. Khaldi, D. Alcaraz-Segura, E. Guirado, Y. Benhammou, A. El Afia, F. Herrera, and S. Tabik, "Timespec4lulc: a global multispectral time series database for training lulc mapping models with machine learning," *Earth System Science Data*, vol. 14, no. 3, pp. 1377–1411, 2022.
- [288] X.-Y. Tong, G.-S. Xia, and X. X. Zhu, "Enabling country-scale land cover mapping with meter-resolution satellite imagery," *arXiv preprint arXiv:2209.00727*, 2022.
- [289] J. Li, X. Huang, and L. Tu, "Whu-ohs: A benchmark dataset for large-scale hyperspectral image classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, p. 103022, 2022.
- [290] N. Johnson, W. Treble, and D. Crispell, "Opensentinelmap: A large-scale land use dataset using openstreetmap and sentinel-2 imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1333–1341.
- [291] B. Le Saux, N. Yokoya, R. Hänsch, and S. Prasad, "2018 ieee grss data fusion contest: Multimodal land use classification [technical committees]," *IEEE geoscience and remote sensing magazine*, vol. 6, no. 1, pp. 52–54, 2018.
- [292] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, and G. Forestier, "Multisenge: a multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 635–640, 2022.
- [293] N. Hurst-Tarrab, L. Chang, M. Gonzalez-Mendoza, and N. Hernandez-Gress, "Robust parking block segmentation from a surveillance camera perspective," *Applied Sciences*, vol. 10, no. 15, p. 5364, 2020.
- [294] S. He and H. Balakrishnan, "Lane-level street map extraction from aerial imagery," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2080–2089.
- [295] R. Hänsch, J. Arndt, D. Lunga, M. Gibb, T. Pedelose, A. Boedihardjo, D. Petrie, and T. M. Bacastow, "Spacenet 8—the detection of flooded roads and buildings," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1472–1480.
- [296] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, and S. L. Waslander, "Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 42–55, 2019.
- [297] Drivendata. (2019) Open ai caribbean challenge:mapping disaster risk from aerial imagery. [Online]. Available: <https://www.drivendata.org/competitions/58/disaster-response-roof-type/>
- [298] S. Krapf, L. Bogenrieder, F. Netzler, G. Balke, and M. Lienkamp, "Rid—roof information dataset for computer vision-based photovoltaic potential assessment," *Remote Sensing*, vol. 14, no. 10, p. 2299, 2022.
- [299] Q. Zhang, R. Cong, C. Li, M.-M. Cheng, Y. Fang, X. Cao, Y. Zhao, and S. Kwong, "Dense attention fluid network for salient object detection in optical remote sensing images," *IEEE Transactions on Image Processing*, vol. 30, pp. 1305–1317, 2020.
- [300] S. Luo, H. Li, and H. Shen, "Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 443–457, 2020.
- [301] (2020) Bh-pools/watertanks datasets. [Online]. Available: <http://patreο.dcc.ufmg.br/2020/07/29/bh-pools-watertanks-datasets/>
- [302] S. M. Azimi, C. Henry, L. Sommer, A. Schumann, and E. Vig, "Skyscapes fine-grained semantic understanding of aerial scenes," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7393–7403.
- [303] R. Abdelfattah, X. Wang, and S. Wang, "Ttpla: An aerial-image dataset for detection and segmentation of transmission towers and power lines," in *Proceedings of the Asian Conference on Computer Vision*, 2020.
- [304] (2020) Satellite images of water bodies. [Online]. Available: <https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies>
- [305] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, "Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [306] A. Fujita, K. Sakurada, T. Imaizumi, R. Ito, S. Hikosaka, and R. Nakamura, "Damage detection from aerial images via convolutional neural networks," in *2017 Fifteenth IAPR international conference on machine vision applications (MVA)*. IEEE, 2017, pp. 5–8.
- [307] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sensing*, vol. 12, no. 10, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/10/1662>
- [308] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston, "xbd: A dataset for assessing building damage from satellite imagery," *arXiv preprint arXiv:1911.09296*, 2019.
- [309] M. Liu, Z. Chai, H. Deng, and R. Liu, "A cnn-transformer network with multi-scale context aggregation for fine-grained cropland change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022.
- [310] C. Benedek and T. Szirányi, "Change detection in optical aerial images by a multilayer conditional mixed markov model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 10, pp. 3416–3430, 2009.
- [311] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2858–2874, 2013.
- [312] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS journal of photogrammetry and remote sensing*, vol. 107, pp. 50–63, 2015.
- [313] Q. Wang, Z. Yuan, Q. Du, and X. Li, "Getnet: A general end-to-end 2-d cnn framework for hyperspectral image change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2018.
- [314] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2018.
- [315] (2018) Aerial change detection in video games. [Online]. Available: <https://www.kaggle.com/datasets/kmader/aerial-change-detection-in-video-games>
- [316] M. Lebedev, Y. V. Vizilter, O. Vygolov, V. Knyaz, and A. Y. Rubis, "Change detection in remote sensing images using conditional adversarial networks," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, no. 2, 2018.
- [317] (2016) Multitemporal hyperspectral change detection. [Online]. Available: <https://gitlab.citius.usc.es/hiperespectral/ChangeDetectionDataset>
- [318] C. Wu, L. Zhang, and L. Zhang, "A scene change detection framework for multi-temporal very high resolution remote sensing images," *Signal Processing*, vol. 124, pp. 184–197, 2016.
- [319] M. Zhang and W. Shi, "A feature difference convolutional neural network-based change detection method," *IEEE Transactions on Geoscience and Remote Sensing*, 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9052762>
- [320] C. Zhang, P. Yue, D. Tapete, L. Jiang, B. Shangguan, L. Huang, and G. Liu, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 183–200, 2020.
- [321] J. López-Fandiño, A. S. Garea, D. B. Heras, and F. Argüello, "Stacked autoencoders for multiclass change detection in hyperspectral images," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2018, pp. 1906–1909.
- [322] S. Tian, A. Ma, Z. Zheng, and Y. Zhong, "Hi-ucd: A large-scale dataset for urban semantic change detection in remote sensing imagery," *arXiv preprint arXiv:2011.03247*, 2020.
- [323] D. Peng, L. Bruzzone, Y. Zhang, H. Guan, H. Ding, and X. Huang, "Semidnet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 5891–5906, 2020.
- [324] A. Moghimi, A. Mohammadzadeh, T. Celik, and M. Amani, "A novel radiometric control set sample selection strategy for relative radiometric normalization of multitemporal satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2503–2519, 2020.

- [325] R. Shao, C. Du, H. Chen, and J. Li, "Sunet: Change detection for heterogeneous remote sensing images from satellite and uav using a dual-channel fully convolution network," *Remote Sensing*, vol. 13, no. 18, p. 3750, 2021.
- [326] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE transactions on geoscience and remote sensing*, vol. 60, pp. 1–16, 2021.
- [327] J. Yuan, L. Ru, S. Wang, and C. Wu, "Wh-mavs: A novel dataset and deep learning benchmark for multiple land use and land cover applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1575–1590, 2022.
- [328] M. Leenstra, D. Marcos, F. Bovolo, and D. Tuia, "Self-supervised pre-training enhances change detection in sentinel-2 imagery," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 578–590.
- [329] H. Li, F. Zhu, X. Zheng, M. Liu, and G. Chen, "Msclunet: A deep learning framework for built-up area change detection integrating multispectral, sar and vhr data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022.
- [330] M. Barekatain, M. Martí, H.-F. Shih, S. Murray, K. Nakayama, Y. Matsuo, and H. Prendinger, "Okutama-action: An aerial view video dataset for concurrent human action detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 28–35.
- [331] H.-W. Jo and W.-K. Lee, "Paddy rice maps south korea (2017-2021)," Jan. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.5845896>
- [332] K. Chanwoo, J. Hyun-Woo, L. Sujong, K. Whijin, Y. Yan, K. Joon, S. Minju, and W.-K. Lee, "Paddy rice labeling sites in south korea (2018)," Jan. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.5846018>
- [333] P. Jin, L. Mou, G.-S. Xia, and X. X. Zhu, "Anomaly detection in aerial videos with transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [334] H. Goldberg, M. Brown, and S. Wang, "A benchmark for building footprint classification using orthorectified rgb imagery and digital surface models from commercial satellites," in *Proceedings of IEEE Applied Imagery Pattern Recognition Workshop 2017*, 2017.
- [335] (2016) Draper satellite image chronology. [Online]. Available: <https://www.kaggle.com/c/draper-satellite-image-chronology>
- [336] R. Bahmanyar, E. Vig, and P. Reinartz, "Mrnet: Crowd counting and density map estimation in aerial and ground imagery," *arXiv preprint arXiv:1909.12743*, 2019.
- [337] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The sen1-2 dataset for deep learning in sar-optical data fusion," *arXiv preprint arXiv:1807.01569*, 2018.
- [338] M. Huang, Y. Xu, L. Qian, W. Shi, Y. Zhang, W. Bao, N. Wang, X. Liu, and X. Xiang, "The qxs-saropt dataset for deep learning in sar-optical data fusion," *arXiv preprint arXiv:2103.08259*, 2021.
- [339] B. Huang, L. Zhi, C. Yang, F. Sun, and Y. Song, "Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, pp. 1806–1813.
- [340] J. M. R. Mullor, "jmrmcode/Lidar-Landsat-data-fusion: lidar-landsat-data-fusion," Oct. 2019. [Online]. Available: <https://doi.org/10.5281/zenodo.3468645>
- [341] Y. Wang, N. Ait Ali Braham, C. M. Albrecht, Z. Xiong, C. Liu, and X. Zhu, "Ssl4eo-s12: A large-scale multimodal multitemporal dataset for self-supervised learning in earth observation," 2022. [Online]. Available: <https://mediatum.ub.tum.de/1660427>
- [342] L. Liu and H. Li, "Lending orientation to neural networks for cross-view geo-localization," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [343] S. Zhu, T. Yang, and C. Chen, "Vigor: Cross-view image geo-localization beyond one-to-one retrieval," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 3640–3649.
- [344] C. Wang, A. Mouche, P. Tandeo, J. E. Stopa, N. Longépé, G. Erhard, R. C. Foster, D. Vandemark, and B. Chapron, "A labelled ocean sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode," *Geoscience Data Journal*, vol. 6, no. 2, pp. 105–115, 2019.
- [345] S. Cavegn, N. Haala, S. Nebiker, M. Rothermel, and P. Tutzauer, "Benchmarking high density image matching for oblique airborne imagery," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, no. 3, p. 45, 2014.
- [346] J. L. H. R.-A. García, Lina; Díaz. (2019) Thermal and visible aerial imagery. [Online]. Available: <https://data.mendeley.com/datasets/ffgxzxz298/1>
- [347] WHU. (2019) Whu-sen-city. [Online]. Available: <https://github.com/whu-csl/WHU-SEN-City>
- [348] G. Facciolo, C. De Franchis, and E. Meinhardt-Holzapfel, "Automatic 3d reconstruction from multi-date satellite images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 57–66.
- [349] J. Gao, J. Liu, and S. Ji, "Rational polynomial camera model warping for deep learning based satellite multi-view stereo matching," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 6148–6157.
- [350] E. López-Jiménez, J. I. Vasquez-Gomez, M. A. Sanchez-Acevedo, J. C. Herrera-Lozada, and A. V. Uriarte-Arcia, "Columnar cactus recognition in aerial images using a deep learning approach," *Ecological Informatics*, vol. 52, pp. 131–138, 2019.
- [351] G. Reiersen, D. Dao, B. Lütjens, K. Klemmer, K. Amara, A. Steinegger, C. Zhang, and X. Zhu, "Reforestree: A dataset for estimating tropical forest carbon stock with deep learning and aerial imagery," *arXiv preprint arXiv:2201.11192*, 2022.
- [352] J. Ventura, M. Honsberger, C. Gonsalves, J. Rice, C. Pawlak, N. L. Love, S. Han, V. Nguyen, K. Sugano, J. Doremus et al., "Individual tree detection in large-scale urban environments using high-resolution multispectral imagery," *arXiv preprint arXiv:2208.10607*, 2022.
- [353] S. Doda, Y. Wang, M. Kahl, E. J. Hoffmann, H. Taubenböck, and X. X. Zhu, "So2sat pop-a curated benchmark data set for population estimation from space on a continental scale," *arXiv preprint arXiv:2204.08524*, 2022.
- [354] T. H. Park, M. Märtens, G. Lecuyer, D. Izzo, and S. D'Amico, "Speed+: Next-generation dataset for spacecraft pose estimation across domain gap," in *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022, pp. 1–15.
- [355] (2020) Satellite images to predict poverty. [Online]. Available: <https://www.kaggle.com/datasets/sandeshbhat/satellite-images-to-predict-povertyafrica>
- [356] S. Lobry, J. Murray, D. Marcos, and D. Tuia, "Visual question answering from remote sensing images," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 4951–4954.
- [357] A. Djerida, K. Djerriri, M. S. Karoui et al., "A new public alsat-2b dataset for single-image super-resolution," in *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 8095–8098.
- [358] J. Cornebise, I. Oršolić, and F. Kalaitzis, "The WorldStrat Dataset: Open High-Resolution Satellite Imagery With Paired Multi-Temporal Low-Resolution," Jul. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.6810792>
- [359] J. Michel, J. Vinasco-Salinas, J. Inglada, and O. Hagolle, "SEN2VENμS, a dataset for the training of Sentinel-2 super-resolution algorithms," May 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.6514159>
- [360] M. Märtens, D. Izzo, A. Krzic, and D. Cox, "Probabilistic super-resolution dataset," Nov. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.6327426>
- [361] M. Brown, H. Goldberg, K. Foster, A. Leichtman, S. Wang, S. Hagstrom, M. Bosch, and S. Almes, "Large-scale public lidar and satellite image data set for urban semantic labeling," in *Laser Radar Technology and Applications XXIII*, vol. 10636. SPIE, 2018, pp. 154–167.
- [362] T. Wu, B. Vallet, M. Pierrot-Deseilligny, and E. Rupnik, "A new stereo dense matching benchmark dataset for deep learning," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 405–412, 2021.
- [363] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*, 2019.
- [364] N. Varney, V. K. Asari, and Q. Graehling, "Dales: A large-scale aerial lidar data set for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 186–187.
- [365] X. Li, G. Cheng, S. Liu, Q. Xiao, M. Ma, R. Jin, T. Che, Q. Liu, W. Wang, Y. Qi et al., "Heihe watershed allied telemetry experi-

- mental research (hiwater): Scientific objectives and experimental design," *Bulletin of the American Meteorological Society*, vol. 94, no. 8, pp. 1145–1160, 2013.
- [366] Q. Hu, B. Yang, S. Khalid, W. Xiao, N. Trigoni, and A. Markham, "Sensaturban: Learning semantics from urban-scale photogrammetric point clouds," *International Journal of Computer Vision*, vol. 130, no. 2, pp. 316–343, 2022.
- [367] W. Gao, L. Nan, B. Boom, and H. Ledoux, "Sum: A benchmark dataset of semantic urban meshes," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 108–120, 2021.
- [368] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert, "Contextual classification with functional max-margin markov networks," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 975–982.
- [369]
- [370] X. Roynard, J.-E. Deschaud, and F. Goulette, "Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification," *The International Journal of Robotics Research*, vol. 37, no. 6, pp. 545–557, 2018.
- [371] S. Zolanvari, S. Ruano, A. Rana, A. Cummins, R. E. da Silva, M. Rahbar, and A. Smolic, "Dublincity: Annotated lidar point cloud and its applications," *arXiv preprint arXiv:1909.03613*, 2019.
- [372] A. Serna, B. Marcotegui, F. Goulette, and J.-E. Deschaud, "Paris-rue-madame database: a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods," in *4th international conference on pattern recognition, applications and methods ICPRAM 2014*, 2014.
- [373] (2022) Aerial monocular 3d object detection dataset. [Online]. Available: <https://sjtu-magic.github.io/dataset/AM3D/>
- [374] S. Wang, J. Li, W. Zhang, C. Cao, F. Zhang, Q. Shen, X. Zhang, and B. Zhang, "A dataset of remote-sensed Forel-Ule Index for global inland waters during 2000–2018," 12 2020. [Online]. Available: https://figshare.com/articles/dataset/A_dataset_of_remote-sensed_Forel-Ule_Index_for_global_inland_waters_during_2000_2018/13014299
- [375] S. Yadav, C. Fu, and A. v. Wyk, (2017) Historical hourly weather data 2012–2017. [Online]. Available: <https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data>
- [376] C. Requena-Mesa, V. Benson, M. Reichstein, J. Runge, and J. Denzler, "Earthnet2021: A large-scale dataset and challenge for earth surface forecasting as a guided video prediction task," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1132–1142.
- [377] (2017) Climate change: Earth surface temperature data. [Online]. Available: <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>
- [378] (2017) International greenhouse gas emissions. [Online]. Available: <https://www.kaggle.com/datasets/unitednations/international-greenhouse-gas-emissions>
- [379] M. Maskey, R. Ramachandran, M. Ramasubramanian, I. Gurung, B. Freitag, A. Kaulfus, D. Bollinger, D. J. Cecil, and J. Miller, "DeepTi: Deep-learning-based tropical cyclone intensity estimation system," *IEEE journal of selected topics in applied Earth observations and remote sensing*, vol. 13, pp. 4271–4281, 2020.
- [380] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [381] I. Dimitrovski, I. Kitanovski, P. Panov, N. Simidjievski, and D. Kocev, "Aitlas: Artificial intelligence toolbox for earth observation," *arXiv preprint arXiv:2201.08789*, 2022.
- [382] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.
- [383] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu *et al.*, "Mmdetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.
- [384] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [385] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.
- [386] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976–11986.
- [387] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [388] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit *et al.*, "Mlp-mixer: An all-mlp architecture for vision," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24261–24272, 2021.
- [389] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.
- [390] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [391] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [392] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
- [393] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified perceptual parsing for scene understanding," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 418–434.
- [394] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12077–12090, 2021.
- [395] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.
- [396] M. Lu, J. Liu, F. Wang, and Y. Xiang, "Multi-task learning of relative height estimation and semantic segmentation from single airborne rgb images," *Remote Sensing*, vol. 14, no. 14, p. 3450, 2022.
- [397] G. Sumbul, J. Kang, T. Kreuziger, F. Marcelino, H. Costa, P. Benedito, M. Caetano, and B. Demir, "Bigearthnet dataset with a new class-nomenclature for remote sensing image understanding," *arXiv preprint arXiv:2001.06372*, 2020.
- [398] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020.
- [399] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [400] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [401] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [402] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.