

IMAGE LICENSED BY INGRAM PUBLISHING

There Are No Data Like More Data

Datasets for deep learning in Earth observation

MICHAEL SCHMITT^{ID}, SEYED ALI AHMADI^{ID}, YONGHAO XU^{ID}, GÜLŞEN TAŞKIN^{ID},
UJJWAL VERMA^{ID}, FRANCESCA PAOLO SICA^{ID}, AND RONNY HÄNSCH^{ID}

Carefully curated and annotated datasets are the foundation of machine learning (ML), with particularly data-hungry deep neural networks forming the core of what is often called *artificial intelligence* (AI). Due to the massive success of deep learning (DL) applied to Earth observation (EO) problems, the focus of the community has been largely on the development of evermore sophisticated deep neural network architectures and training strategies. For that purpose, numerous task-specific datasets have been created that were largely ignored by previously published review articles on AI for EO. With this article, we want to change the perspective and put ML datasets dedicated to EO data and applications into

the spotlight. Based on a review of historical developments, currently available resources are described and a perspective for future developments is formed. We hope to contribute to an understanding that the nature of our data is what distinguishes the EO community from many other communities that apply DL techniques to image data, and that a detailed understanding of EO data peculiarities is among the core competencies of our discipline.

INTRODUCTION

DL techniques have enabled drastic improvements in many scientific fields, especially in those dedicated to the analysis of image data, e.g., computer vision or remote sensing. Although it was possible to train *shallow* learning approaches on comparably small datasets, *deep* learning requires

Digital Object Identifier 10.1109/MGRS.2023.3293459
Date of current version: 8 August 2023

large-scale datasets to reach the desired accuracy and generalization performance. Therefore, the availability of annotated datasets has become a dominating factor for many cases of modern EO data analysis that develops and evaluates powerful, DL-based techniques for the automated interpretation of remote sensing data.

The main goal of general computer vision is the analysis of optical images, such as photos, which contain everyday objects, e.g., furniture, animals, or road signs. Remote sensing involves a larger variety of sensor modalities and image analysis tasks than conventional computer vision, rendering the annotation of remote sensing data more difficult and costly. Besides classical optical images, multi- or hyperspectral sensors and different kinds of infrared sensors; active sensor technologies such as laser scanning, microwave altimeters, and synthetic aperture radar (SAR) are regularly used, too. The fields of application range from computer vision-like tasks, such as object detection and classification, to semantic segmentation (mainly for land cover mapping) to specialized regression tasks grounded in the physics of the used remote sensing system.

To provide an illustrative example, a dataset for biomass regression from interferometric SAR data will adopt imagery and annotations very different from the ones needed for the semantic segmentation of urban land cover types from multispectral optical data. Thus, although extensive image databases, such as ImageNet, were created more than 10 years ago and form the backbone of many modern ML developments in computer vision, there is still no similar dataset or backbone network in remote sensing. (Note that, as a prime example of an annotated computer vision dataset, ImageNet contains more than 14 million images depicting objects from more than 20,000 categories.) This lack of generality renders the generation of an ImageNet-like general EO dataset extremely complicated and thus costly: instead of photographs openly accessible on the Internet, many different—and sometimes quite expensive—sensor data would have to be acquired and, instead of “mechanical turks,” trained EO experts would have to be hired to link these different sensor data to the multitude of different domain- and task-specific annotations (see “The Mechanical Turk”).

Therefore, until now, the trend in ML applied to EO data has been characterized by the generation of

numerous remote sensing datasets, each consisting of a particular combination of sensor modalities, applications, and geographic locations. Yet, a review of these developments is still missing in the literature. The only articles that make a small step toward a general review of benchmark datasets are [1], [2], [3], and [4]. All of them provide some sort of review, however, they are always limited to a very narrow aspect, e.g., object detection or scene classification. Furthermore, their focus is on ML approaches and their corresponding datasets, while the historical evolution of datasets is neither discussed in detail, nor from a sensor- and task-agnostic point of view.

As an extension of our 2021 IEEE International Symposium on Geoscience and Remote Sensing contribution [5], this article intends to close this gap by

- ▷ reviewing current developments in the creation of datasets for DL applications in remote sensing and EO
- ▷ structuring existing datasets and discussing their properties
- ▷ providing a perspective on future requirements.

In this context, we additionally present the Earth Observation Database (EOD) [6], which is the result of the effort and cooperation of voluntary scientists within the IEEE Geoscience and Remote Sensing Society (GRSS) Image Analysis and Data Fusion (IADF) Technical Committee (TC). This database aims to function as a centralized tool that organizes the meta information about existing datasets in a community-driven manner.

EVOLUTION OF EO-ORIENTED ML DATASETS

HISTORICAL DEVELOPMENT

High-quality benchmark datasets have played an increasingly important role in EO for quite some time and are one of the driving factors for the recent success of DL approaches to analyze remote sensing data. As such, they can be seen as a tool complementary to methodological advancements to push accuracy, robustness, and generalizability. This section reviews and summarizes the historical development of EO-oriented ML datasets to provide insights into the evolution of this “tool,” ranging from its historical beginnings to the current state of the art.

The beginnings of ML applied to remote sensing focused on specific applications. Datasets were mainly built by considering a very localized study site, a few specific sensor modalities, and a relatively small number of acquired samples. Therefore, the first datasets were relatively small compared to what is now considered a benchmarking dataset. Training, validation, and testing samples were often taken from the same image. Even with the use of sophisticated shallow learning models but especially since the advent of DL, such small datasets were no longer sufficient for proper training and evaluation. The need for extended datasets has led to the creation of larger datasets containing multiple images, often acquired at different geographic locations.

The Mechanical Turk

The name *mechanical turk* comes from a fraudulent chess-playing machine developed in the 18th century. Chess players were made to believe they played against the machine, but were in fact competing against a person hidden inside it. Today, the term mostly refers to Amazon Mechanical Turk (MTurk), a crowdsourcing website run by Amazon. On MTurk, users can hire remotely located crowdworkers to perform desired tasks. MTurk is frequently used to create manual annotations for supervised machine learning tasks.

Figure 1 illustrates the evolution of benchmark datasets for ML in EO by showing this temporal development. To provide a broad view about the recent evolution of

benchmarking datasets, we gathered an extensive list of datasets available in the EO community, resulting in a large collection with currently 400 (i.e., 380 image-based

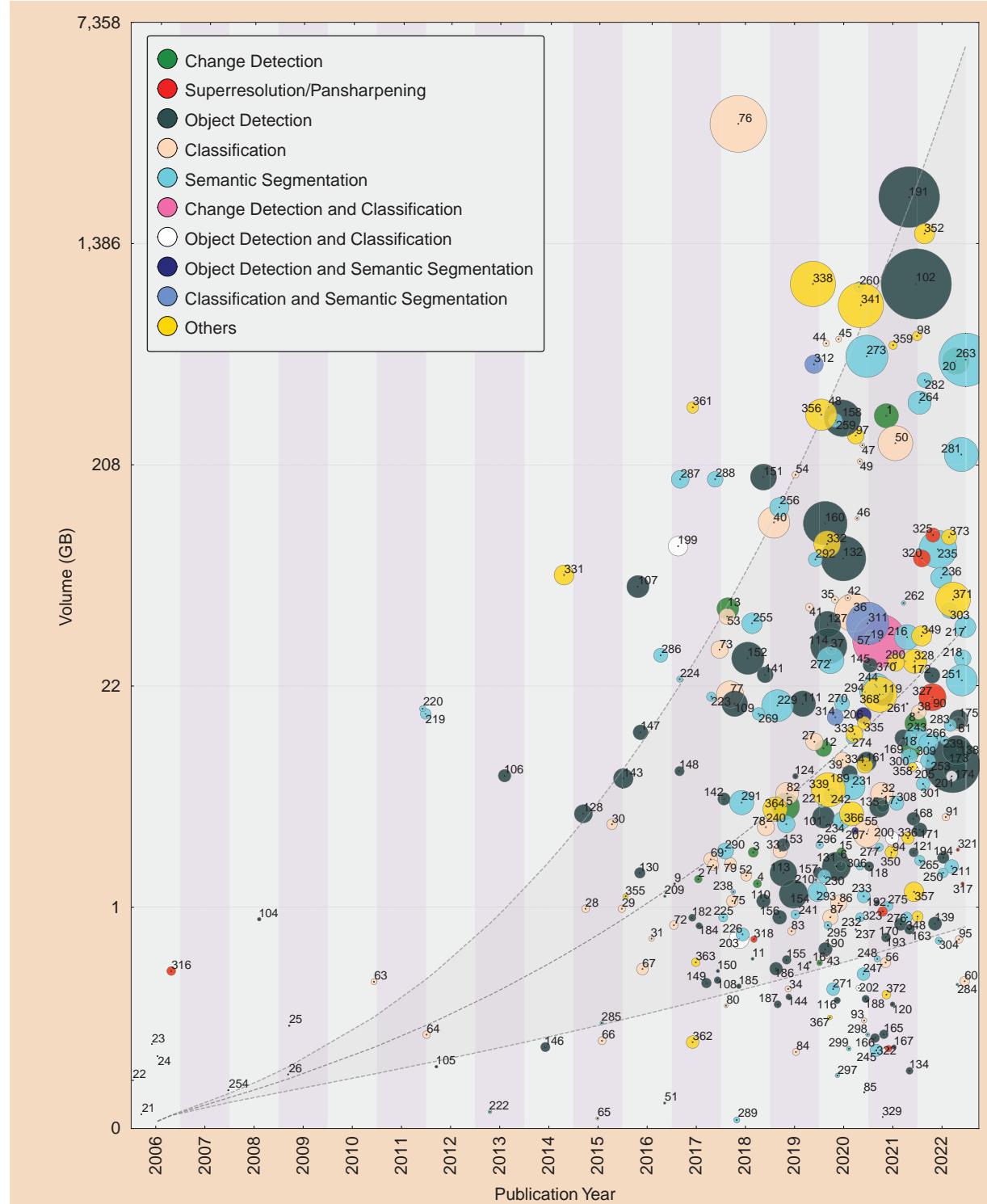


FIGURE 1. Distribution of remote sensing datasets over the years. The x-axis shows the publication year (the datasets are placed within the region of their publication year with a small random offset to minimize visual overlap in the graph), while the y-axis represents the volume of each dataset in gigabytes on a logarithmic scale. The circle radius indicates the dataset size in terms of the number of pixels. Colors denote the type of task addressed by a dataset. Each circle is accompanied by an index, allowing for identification of the dataset in the database (see Table 2), which provides further information.

and 20 point cloud-based datasets) entries (see Tables 1 and 2), including related metadata. We point out that, although being extensive, this list is far from being complete due to the fact that a large number of new datasets are published every year. Furthermore, the metadata required to generate the plot of Figure 1 are available only for a subset of 290 datasets (roughly 73%). In the horizontal axis, we indicate the year of publication. The vertical axis shows the volume of a dataset, while the circle radius reflects the number of spatial pixels covered by a dataset. For a more detailed explanation of how we measure dataset size,

please refer to Figure S1 and “How to Measure the Size of a Dataset.”

Figure 1 provides a straightforward overview of the proportion among size and spatial dimension, and therefore about the overall information content given by features such as resolution, sensor modalities, number of bands/channels, and so on. Each circle is accompanied by an index, allowing for identification of the dataset in the database (see Table 2) that provides further information. Note that we use the category “Others” for datasets that do not belong to any of the other categories and are too rare to

TABLE 1. ALTHOUGH A THOROUGH ANALYSIS OF LIDAR DATASETS IS BEYOND THE SCOPE OF THIS SURVEY, WE DO PROVIDE AN OVERVIEW OF SEVERAL EXAMPLE DATASETS. POINT CLOUD DATASETS ARE ANOTHER LARGE GROUP OF BENCHMARK DATA THAT ARE WIDELY USED IN THE LITERATURE AND INDUSTRY. WITHIN EO THE MOST COMMON SOURCE FOR POINT CLOUD DATA ARE LIDAR SENSORS THAT USE LIGHT IN THE FORM OF LASER PULSES TO MEASURE THE DISTANCE TO THE SURFACE. THE PRIMARY SOURCES ARE AIRBORNE LASER SCANNING (ALS), TERRESTRIAL LASER SCANNING (TLS), AND MOBILE LASER SCANNING (MLS) DEVICES. OTHER SOURCES OF POINT CLOUDS AND 3D DATA INCLUDE PHOTOGRAHMETRIC METHODS (STRUCTURE FROM MOTION, MULTI-VIEW STEREO, AND DENSE MATCHING APPROACHES) AND TOMOGRAPHIC SAR. AS 3D DATA TYPICALLY COME WITH FEATURES THAT ARE VERY DIFFERENT FROM 2D IMAGE DATA, SUCH DATASETS ARE BEYOND THE SCOPE OF THIS ARTICLE. NEVERTHELESS, TABLE 1 PROVIDES A SHORT LIST OF EXAMPLE LIDAR/POINT CLOUD DATASETS FOR INTERESTED READERS.

TASK	PLATFORM	TIMESTAMPS	NAME	PUBLICATION DATE	POINT DENSITY (POINTS/M ²)	NUMBER OF CLASSES	NUMBER OF POINTS	VOLUME (MB)
Change detection	ALS	Multiple	Abenberg ALS	2013	16	—	5,400,000	258
Classification	ALS	Single	NEWFOR	2015	Varies	Four	—	97
Classification	ALS	Single	DFC19	2019	—	Six	167,400,000	613
Classification	ALS	Single	ISPRS 3D Vaihingen	2014	8	Nine	780,879	—
Classification	Multiple	Single	ArCH	2020	Varies	10	136,138,423	—
Classification/semantic segmentation	ALS	Single	DublinCity	2019	240–348	13	260,000,000	3,000
Filtering	ALS	Single	OpenGF	2021	6 and 14	Three	542,100,000	2,280
Object detection/semantic segmentation	TLS	Single	LiSurveying	2021	Varies	54	2,450,000,000	—
Others	ALS	Single	RoofN3D	2018	4.72	Three	118,100	—
Semantic segmentation	ALS	Single	LASDU	2020	3–4	Six	3,120,000	—
Semantic segmentation	ALS	Single	DALES	2020	50	Eight	505,000,000	4,000
Semantic segmentation	ALS	Single	DALES Object	2021	50	Eight	492,000,000	5,000
Semantic segmentation	Drone	Single	Campus3D	2020	Varies	24	937,000,000	2,500
Semantic segmentation	Drone	Multiple	Hessigheim 3D	2021	800	11	73,909,354	5,950
Semantic segmentation	Drone	Single	WildForest3D	2022	60	Six	7,000,000	81
Semantic segmentation	MLS	Single	Toronto3D	2020	1,000	Eight	7,830,0000	1,100
Semantic segmentation	MLS	Multiple	HelixNet	2022	—	Nine	8,850,000,000	235,700
Semantic segmentation	Photogrammetry	Single	SensatUrban	2020	—	13	2,847,000,000	36,000
Semantic segmentation	Photogrammetry	Single	STPLS3D	2022	—	20	—	36,600 (images: 700,000)
Semantic segmentation	TLS	Single	Semantic 3D	2017	—	Eight	4000000000	23,940

TABLE 2. ALTHOUGH THE COMPLETE COLLECTION OF 380 DATASETS IS TOO EXTENSIVE TO BE INCLUDED IN THE PRINT VERSION OF THIS ARTICLE, IT CAN BE FOUND AT [85]. HERE WE INCLUDE A PORTION OF THIS LIST, WHICH CONTAINS ALL THE DATASETS MENTIONED IN THE ARTICLE, ALONG WITH THEIR INDEX IN THE COMPLETE DATABASE.

INDEX	TASK	PLATFORM	SENSOR TYPE	NAME	PUBLICATION DATE	NUMBER OF IMAGES	IMAGE SIZE	NUMBER OF CLASSES	VOLUME (MB)
1	Change detection	Aerial	Multiple	DFC21-MSD	2021	2,250	4,000	36,000,000,000	325,000
10	Change detection	Satellite	Multiple	DFC09	2009	Multiple	Two	98	—
11	Change detection	Satellite	Multispectral	OneraCD	2018	Multiple	24	600	8,640,000
15	Change detection	Satellite	Optical	LEVIR-CD	2020	Single	637	1,024	667,942,912
21	Classification	Aerial	Hyperspectral	Indian Pines	2000	Single	One	145	21,025
22	Classification	Aerial	Hyperspectral	Salinas	2000	Single	One	365	111,104
24	Classification	Aerial	Hyperspectral	Kennedy Space Center	2005	Single	One	550	311,100
34	Classification	Drone	Optical	ALIDER	2019	Single	2,645	240	152,352,000
40	Classification	Satellite	Multiple	BigEarthNet-MM	2019	Single	590,326 × 12	120	1,02E + 11
52	Classification	Satellite	Multispectral	EuroSAT	2018	Single	27,000 × 13	64	1,437,696,000
63	Classification	Satellite	Optical	University of California, Merced	2010	Single	2,100	256	137,625,600
69	Classification	Satellite	Optical	AIID	2017	Single	10,000	600	3,600,000,000
76	Classification	Satellite	Optical	FMoW	2018	Single	523,846	—	1,08E + 12
98	Cloud removal	Satellite	Multiple	SEN12MS-CR-TS	2021	Multiple	53	4,000	848,000,000
105	Object detection	Aerial	Optical	SZTAKI AirChange	2012	Multiple	13	800	7920640
109	Object detection	Aerial	Optical	DOTA v1.0	2018	Single	2,806	4,000	44,896,000,000
114	Object detection	Aerial	Optical	DOTA v2.0	2020	Single	11,268	4,000	1,80E + 11
117	Object detection	Aerial	Optical	RSOC	2020	Single	3,057	2,500	3,621,481,392
127	Object detection	Drone	Multispectral	BIRDSAI	2020	Multiple	162,000	640	49,766,400,000
138	Object detection	Drone	Optical	DroneCrowd	2022	Multiple	33,600	1,920	69,672,960,000
151	Object detection	Satellite	Optical	SpaceNet-4	2018	Single	60,000	900	48,600,000,000
162	Object detection	Satellite	Optical	AI-TOD	2020	Single	28,036	—	186,000
191	Object detection/ classification	Satellite	SAR	xView3-SAR	2021	Single	991 × 2	27,000	—
204	Object detection/ classification	Satellite	SAR	FUSAR-Ship	2020	Single	5,000	512	1,310,720,000
208	Semantic segmentation	Aerial	Hyperspectral	DFC08	2008	Single	Five	194	—
211	Semantic segmentation	Aerial	Hyperspectral	HOISD	2022	Single	18 × 224	1,700	3,622,355,072
212	Semantic segmentation	Aerial	Multiple	DFC13	2013	Single	145	16	One
213	Semantic segmentation	Aerial	Multiple	DFC14	2014	Single	Seven	—	—
214	Semantic segmentation	Aerial	Multiple	DFC15	2015	Single	Seven	—	—
215	Semantic segmentation	Aerial	Multiple	DFC18	2018	Single	One	20	19,763

(Continued)

TABLE 2. ALTHOUGH THE COMPLETE COLLECTION OF 380 DATASETS IS TOO EXTENSIVE TO BE INCLUDED IN THE PRINT VERSION OF THIS ARTICLE, IT CAN BE FOUND AT [85]. HERE WE INCLUDE A PORTION OF THIS LIST, WHICH CONTAINS ALL THE DATASETS MENTIONED IN THE ARTICLE, ALONG WITH THEIR INDEX IN THE COMPLETE DATABASE. (Continued)

INDEX	TASK	PLATFORM	SENSOR TYPE	NAME	PUBLICATION DATE	NUMBER OF IMAGES	IMAGE SIZE	SIZE	NUMBER OF CLASSES	VOLUME (MB)	
217	Semantic segmentation	Aerial	Multiple	DFC22-SSL	2022	Single	2,000	20,000,000,000	14	42,000	
219	Semantic segmentation	Aerial	Multispectral	ISPRS 2 D-Potsdam	2011	Single	5,000	1,368,000,000	Six	16,000	
220	Semantic segmentation	Aerial	Multispectral	ISPRS 2 D-Vaihingen	2011	Single	38	6,000	1,368,000,000	Six	17,000
246	Semantic segmentation	Multiple	Multiple	DFC17	2017	Multiple	33	2,200	156,750,000	Six	—
247	Semantic segmentation	Multiple	Multiple	SpaceNet-6	2020	Single	17	—	—	—	
253	Semantic segmentation	Multiple	Optical	OpenEarthMap	2022	Single	3,401	900	275,481,000	One	368
254	Semantic segmentation	Satellite	Multiple	DFC07	2007	Single	5,000	1,024	5,242,880,000	Eight	9,100
257	Semantic segmentation	Satellite	Multiple	DFC20	2020	Single	One	787	619,369	19	—
261	Semantic segmentation	Satellite	Multiple	DFC21-DSE	2021	Single	180,662	17	19,200	—	—
264	Semantic segmentation	Satellite	Multiple	MapInWild	2022	Single	6,000	16	1,536,000	Four	18,000
282	Semantic segmentation	Satellite	Multispectral	OpenSentinelMap	2022	Multiple	1,018 × 8	1,920	30,022,041,600	11	365,000
286	Semantic segmentation	Satellite	Optical	SpaceNet-1	2016	Single	137,045	192	5,052,026,880	15	445,000
287	Semantic segmentation	Satellite	Optical	SpaceNet-2	2017	Single	9,735	650	10,387,585,000	One	31,000
288	Semantic segmentation	Satellite	Optical	SpaceNet-3	2017	Single	24,586	650	6,271,590,000	One	182,200
292	Semantic segmentation	Satellite	Optical	SpaceNet-5	2019	Single	3,711	1,300	4,003,610,000	One	182,200
294	Semantic segmentation	Satellite	Optical	SpaceNet-7	2020	Multiple	2,369	1,300	4,113,037,500	One	84,103
301	Semantic segmentation	Satellite	Optical	SpaceNet-8	2022	Multiple	1,525	1,024	1,599,078,400	One	20,582
312	Semantic segmentation, classification	Satellite	Multiple	SEN12MS	2019	Single	1,200	1,200	3,456,000,000	Two	6,800
313	Semantic segmentation, classification	Satellite	Multiple	DFC23	2023	Single	180,662	256	11,839,864,832	—	510,000
316	Superresolution, pansharpening	Aerial	Multispectral	DFC06	2006	Single	12	—	—	—	—
318	Superresolution, pansharpening	Satellite	Multispectral	Proba-V Superresolution	2018	Single	Six	5,000	637,500,000	—	390
319	Superresolution, pansharpening	Satellite	Multispectral	PAirMax (Airbus)	2021	Single	1,160	384	171,048,960	—	692
324	Superresolution, pansharpening	Satellite	Optical	PAirMax (Maxar)	2021	Single	14	—	—	—	153
325	Superresolution, pansharpening	Satellite	Optical	WorldStrat	2022	Single	3,928	1,054	4,363,678,048	—	386
340	Others	Multiple	DFC19	2019	Single	—	—	—	—	—	107,000
360	Others	Satellite	Optical	DFC16	2016	Multiple	—	—	—	—	—

form their own category. Examples in the “Others” category are datasets on cloud removal, visual question answering, and parameter-estimation tasks such as hurricane wind speed prediction, satellite pose estimation, and vegetation phenological change monitoring. The dashed line (“—”) illustrates an exponential growth of benchmark datasets created by and for the EO community.

This map on the evolution of remote sensing datasets offers several interesting insights:

- **The beginnings:** In addition to the first IEEE GRSS Data Fusion Contest in 2006 (Table 2; number 316), there are a few other pioneering datasets that have fostered ML research applied to remote sensing data in its early stages, e.g.,
 - *Hyperspectral datasets* [(*Indian Pines*, *Salinas Valley*, and *Kennedy Space Center*) (Table 2; 21, 22, and 24)]: Published before 2005, these datasets triggered the ML era in remote sensing. Covering a very small area on the ground and having a very small number of pixels, such datasets are not suitable for training DL models (or have to be used with excessive caution). On the other hand, due to their rich hyperspectral information, they are still being used for tasks such as dimensionality reduction and feature extraction.
 - *The University of California, Merced dataset* (Table 2; 63) [7]: Published in 2010, it was the first dataset dedicated to scene classification.
 - *ISPRS Potsdam/Vaihingen dataset* (Table 2; 219 and 220) [8]: Published in 2012, it was initially intended to benchmark semantic segmentation approaches tailored to aerial imagery. Later, it was also used for other tasks, e.g., single-image height reconstruction (e.g., in [9]).
 - *SZTAKI-AirChange dataset* (Table 2; 105) [10]: Published in 2011, it was one of the earliest datasets designed for object detection.

All of those pioneering datasets have seen massive use in the early ML-oriented EO literature. It is interesting to note that pansharpening, scene classification, semantic segmentation, and object detection were the first topics in remote sensing to be addressed using ML-based methodologies.

- **The DL boom:** As discussed by several review articles on DL and AI applied to EO [11], [12], [13], the year 2015 marked the beginning of the DL boom in the EO community. This is well reflected by a significantly rising number of datasets published from that year onward. It is furthermore confirmed by the fact that the dataset sizes, both in terms of spatial pixels and data volume, started to increase significantly from approximately that time.
- **The diversity of tasks:** From the early days to the present, ML-oriented EO datasets have been designed for a multitude of different tasks. The historical evolution depicted in Figure 1 further shows that object detection and semantic segmentation are the most popular tasks, with a significant increase of datasets dedicated to minority categories (denoted as “Others”) from roughly 2019 to the present. This indicates that the rise of DL in EO broadens the overall application scope of the discipline.

Figure 2 provides further insights into the size distribution of available datasets in terms of the 1) number of pixels and 2) data volume in gigabytes for typical remote sensing tasks. Illustrating these two different aspects allows a deeper understanding of the nature

How to Measure the Size of a Dataset

In this article, we look at the size of datasets from the following two perspectives:

- 1) **Size:** data volume in terms of the number of spatial pixels. We count the number of pixels in the highest-available image resolution while ignoring multiband, multichannel, and multisensor data. In other words, pixels are only counted once in the spatial coverage provided by the dataset.
- 2) **Volume:** data volume in terms of storage. The amount of disk space required for a dataset is a proxy for image resolution and the provided modalities (e.g., multiple bands and sensor types).

Figure S1 highlights the different factors that affect the volume and size of a dataset: the number of bits per pixel (radiometric resolution), number of spectral bands (spectral resolution; i.e., red, green, blue; multispectral; or hyperspectral), number of images during a specific time period (temporal resolution), and number of pixels per unit area (spatial resolution). As mentioned previously, the size is directly related to the unique number of ground-projected resolution cells. A larger dataset in terms of size corresponds to images with higher resolutions or broader coverage.

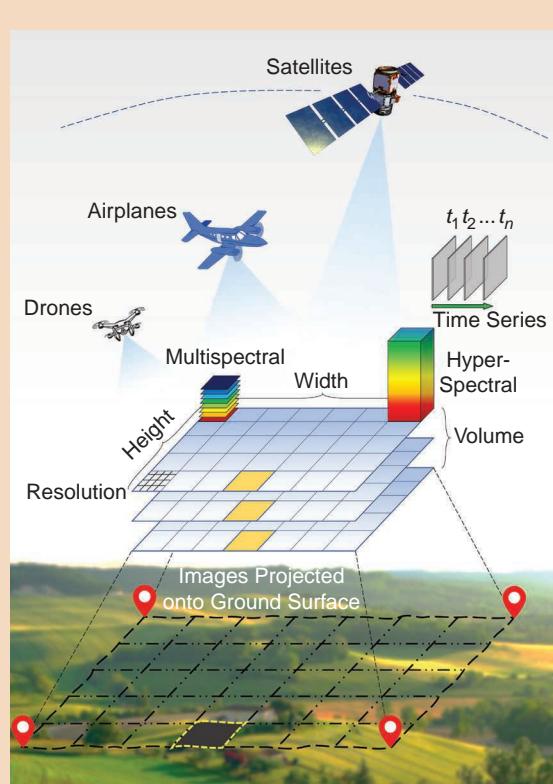


FIGURE S1. A schematic illustration of the proposed size measure used to characterize datasets i.e., pixels are only counted once in the spatial coverage provided by a dataset. For a more detailed definition, see “How to Measure the Size of a Dataset.”

of the data. For example, datasets counting a similar number of spatial pixels may differ in data volume, which can, for example, indicate the use of multimodal imagery. Object detection offers the largest data volume among the existing benchmarking datasets, which

again confirms its popularity in DL-oriented remote sensing research. However, in terms of number of pixels, semantic segmentation takes the lead, indicating that a larger spatial coverage is usually involved for this type of application.

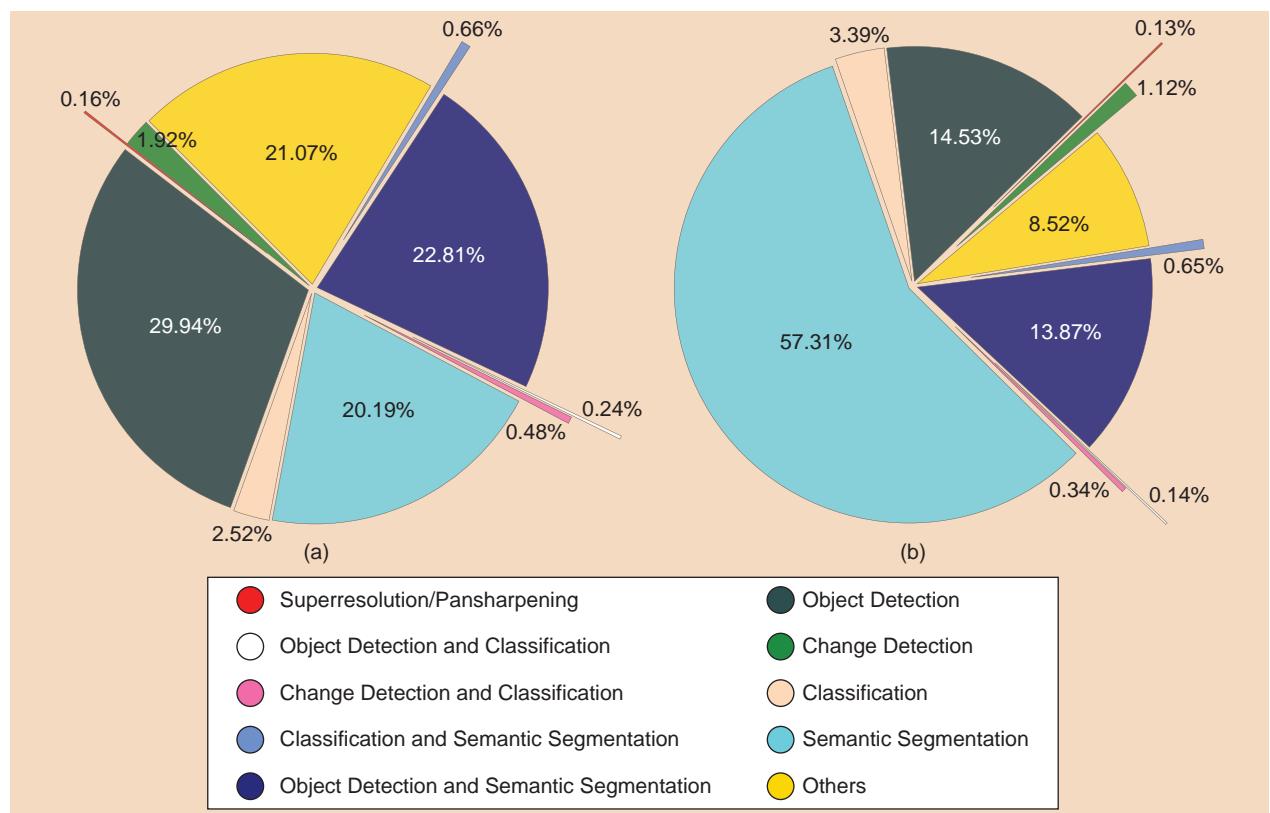


FIGURE 2. Distribution of EO dataset sizes over typical remote sensing tasks, expressed in (a) volume and (b) size, as defined in Figure S1 and “How to Measure the Size of a Dataset.” In (a), with 30%, object detection is the predominant task, followed by semantic segmentation. In (b), in contrast to Figure 2(a), semantic segmentation is the prevailing task, illustrating that corresponding datasets involve more complex scenarios such as leveraging multiple sensors or spectral bands. Object detection and semantic segmentation are the dominant image analysis tasks in ML-centered EO.

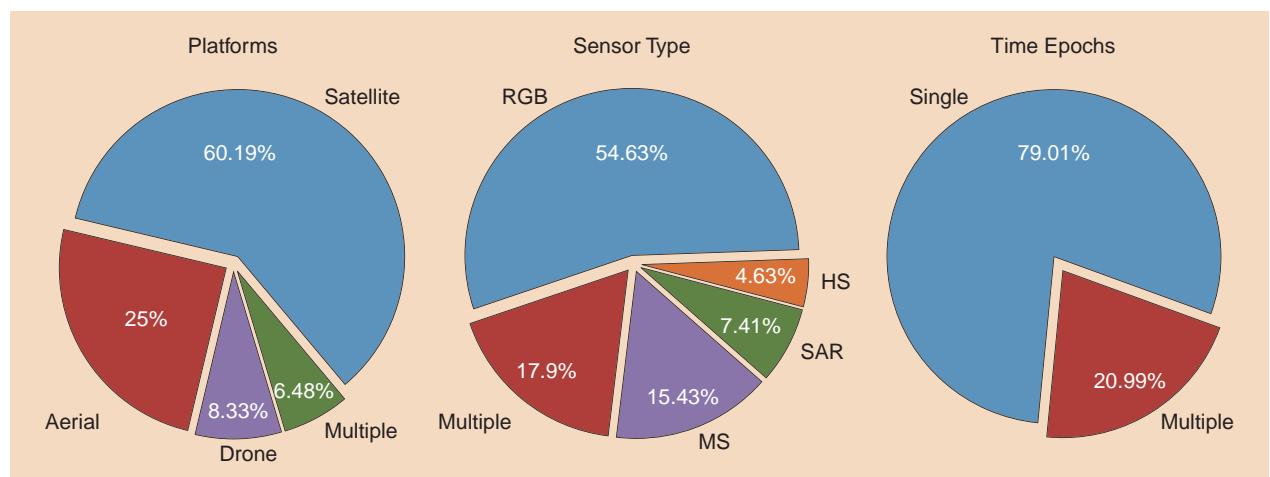


FIGURE 3. A distribution of available EO datasets over different platforms, sensor types, and number of acquisition times. Single-image red, green, blue (RGB) images acquired by satellites are clearly the dominating modality. MS: multispectral; HS: hyperspectral.

PLATFORMS

A direct overview of the occurrence of a type of platform and sensor is given in Figure 3. Satellite platforms are the most common, followed by airborne platforms and drones. Optical data cover more than half of the datasets, while all other sensors are almost equally distributed. Interestingly, 20% of the datasets provide time series, while the rest are single-temporal acquisitions. Complementing this, Figure 4 highlights the distribution of tasks between sensors and platforms. The inner ring indicates

the platform type, which then splits into different sensor types in the middle ring, and finally denotes the targeted tasks in the outer ring, respectively. This graph shows that the datasets acquired by unmanned aerial vehicles (UAVs) and aircraft are mainly dedicated to optical sensors, while satellite-based EO has a much wider and more homogeneous distribution across all sensor and application types.

Figure 5(a) and (b) further specifies the previous findings by showing how the datasets acquired by different platforms are distributed across both tasks and sensors.

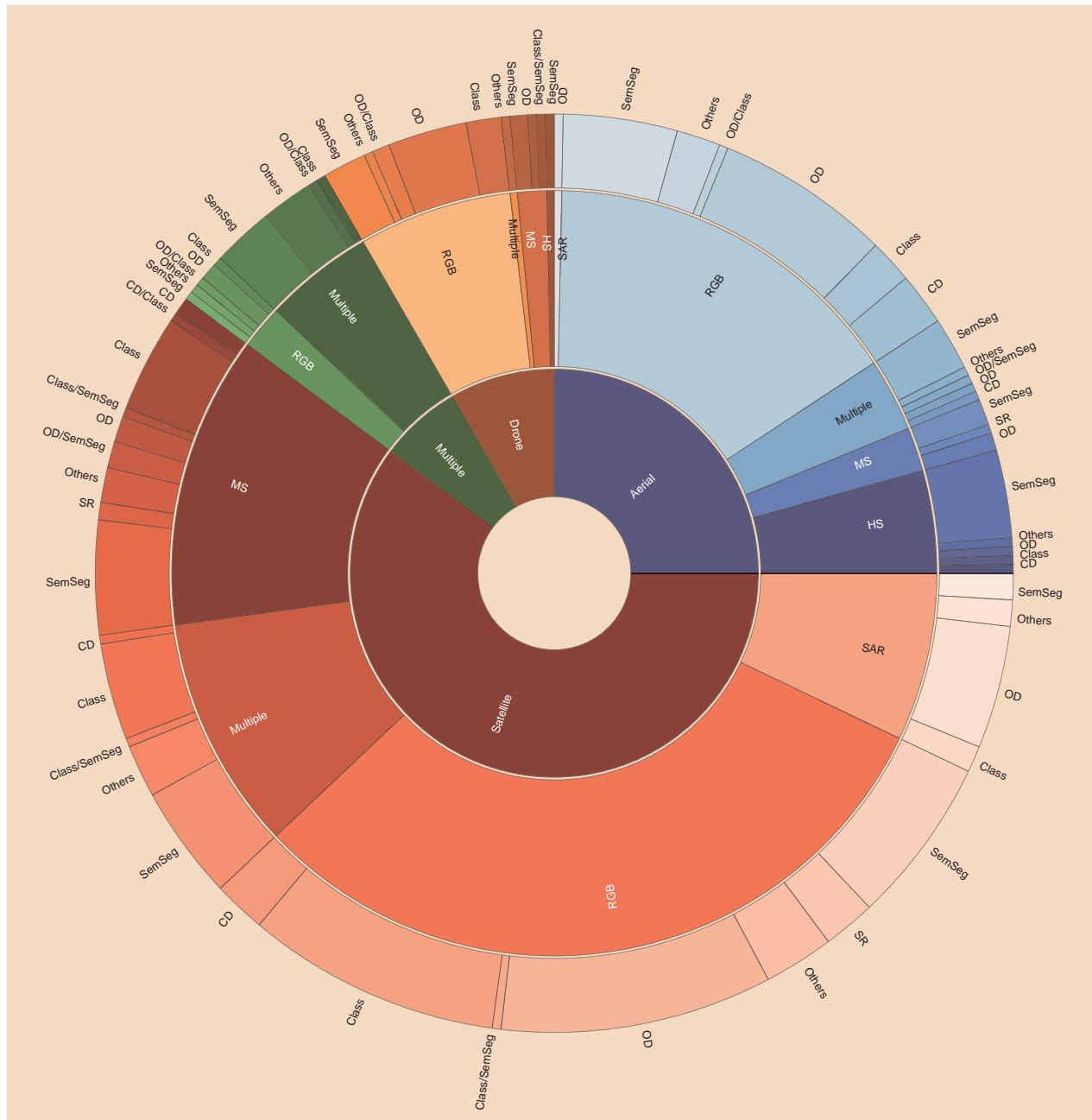


FIGURE 4. A distribution of tasks between sensors and platforms. Platforms are in the inner ring, sensors are distributed in the middle ring, and the outer ring shows different tasks per sensor. OD: object detection; CD: change detection; SemSeg: semantic segmentation; SR: super-resolution; class: classification; MS: multispectral; HS: hyperspectral.

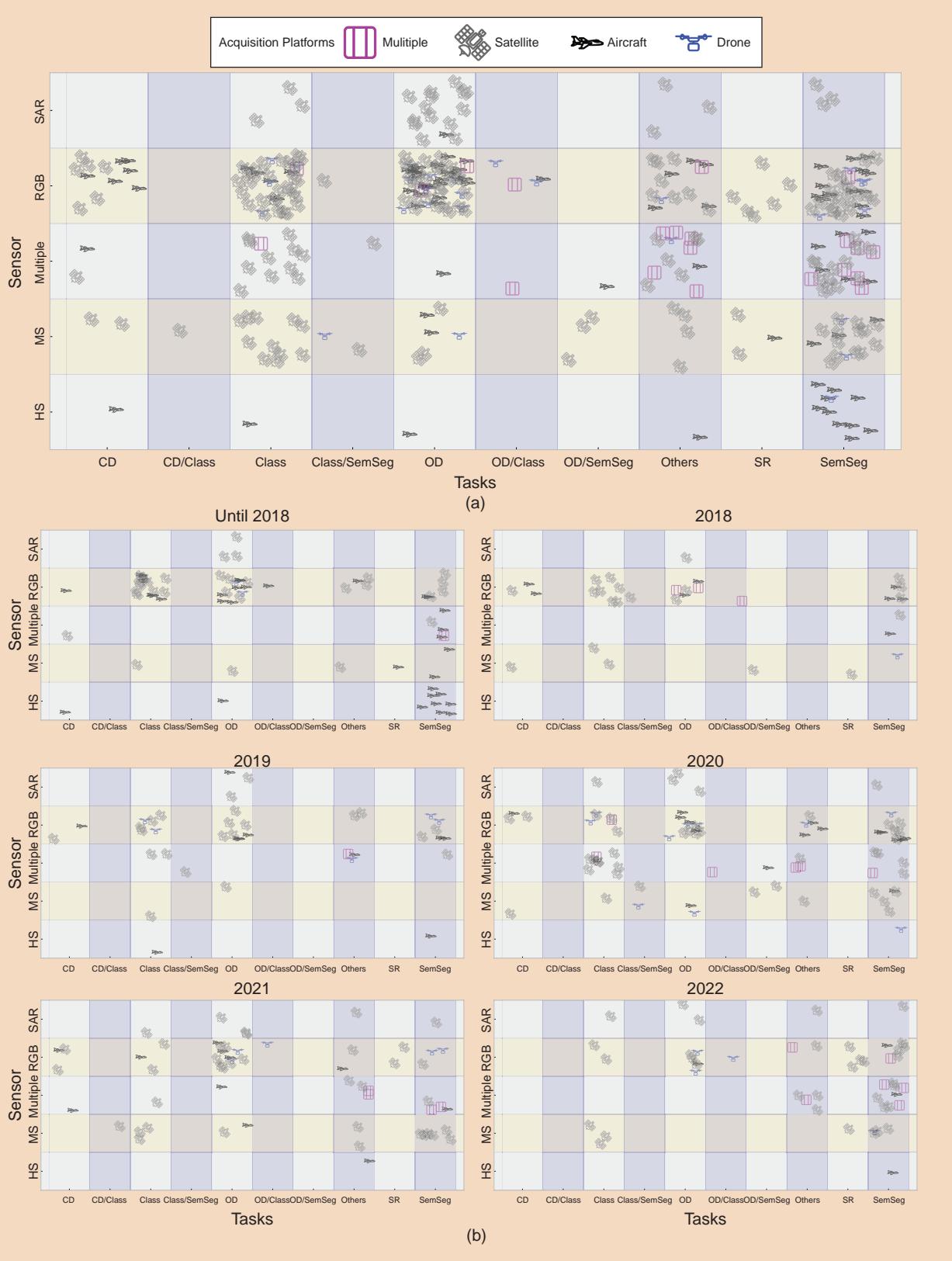


FIGURE 5. Different tasks [CD, classification (class), semantic segmentation (SemSeg), object detection (OD), superresolution (SR), and pansharpening] and task combinations (denoted by "/") make use of very different platforms and sensors. (a) Combinations of different platforms, sensors, and tasks accumulated over the years. (b) Combinations of different platforms, sensors, and tasks for different years. MS: multispectral; HS: hyperspectral.

Although the former provides an overview, the latter adds a temporal aspect by showing the corresponding applications per year. The *x*- and *y*-axes represent tasks and sensors, respectively, while the different markers indicate the type of acquisition platforms. From these plots, we can affirm that object detection and classification tasks are mainly performed on optical images. At the same time, semantic segmentation is fairly evenly distributed among optical, multispectral, and other sensor combinations. SAR images are mainly acquired from satellite platforms, while hyperspectral datasets are almost always acquired from airborne systems. UAVs mainly carry optical sensors in the context of semantic segmentation. Some tasks, such as superresolution, naturally make use of multimodal data, e.g., optical and hyperspectral imagery. The year-by-year graph in Figure 5(b) shows that superresolution datasets, together with UAV-based acquisitions, have received more attention in recent years. On the other hand, the EO community has not seen many new hyperspectral datasets since 2018. Optical sensors were the most common source of information, while after 2020 an increasing number of datasets were also acquired from other sensors, such as SAR or hyperspectral systems.

Figures 3, 4, and 5(a) show that the number of datasets acquired from “Multiple” platforms or sensors is still the minority, which provides evidence for the earlier statement that state-of-the-art datasets are usually designed to respond to a specific task in EO applications. These figures also show which combination of EO tasks, platforms, and sensors is currently underrepresented. In particular, Figure 5(a) shows three main gaps: 1) SAR change detection (CD), 2) SAR superresolution, and 3) hyperspectral superresolution. From a sensor perspective alone, the lack of airborne SAR datasets and drone-based hyperspectral benchmarks are other obvious gaps.

GEOGRAPHIC DIVERSITY

The geographic diversity and spatial distribution of EO imagery is an important attribute of a benchmark dataset as it is directly related to the geographic generalization performance of data-driven models. Figure 6 shows the geographic distribution of datasets (i.e., of roughly ~300 datasets (75%) from Table 2 that provided (or where we could find) their geographic information). Many of the datasets are globally distributed (“Global”) and contain images from around the world, while others cover a limited set of multiple cities or countries (“Multiple Locations”). Maybe surprisingly, “Synthetic” datasets show a dominant presence as well, illustrating the benefits of being able to control image acquisition parameters (such as viewing angle), environmental factors (such as atmospheric conditions, illumination, and cloud cover), and scene content (e.g., type, size, shape, and number of objects). Figure 6 illustrates an important, and seldom-discussed issue within the EO community: there exists a strong geographic bias within the available EO datasets. Although 25% of the datasets contain samples from globally distributed locations, nearly 40% of available datasets cover regions in Europe (21%) and North America (18%) only. Asia is still covered by 10%, however, Africa (5%), South America (4%), and Australia (1%) are barely included. This raises the question of whether many of the findings and conclusions in corresponding research articles would generalize to these geographic areas. In any case, the need for more spatially

OBJECT DETECTION OFFERS THE LARGEST DATA VOLUME AMONG THE EXISTING BENCHMARKING DATASETS, WHICH AGAIN CONFIRMS ITS POPULARITY IN DL-ORIENTED REMOTE SENSING RESEARCH.

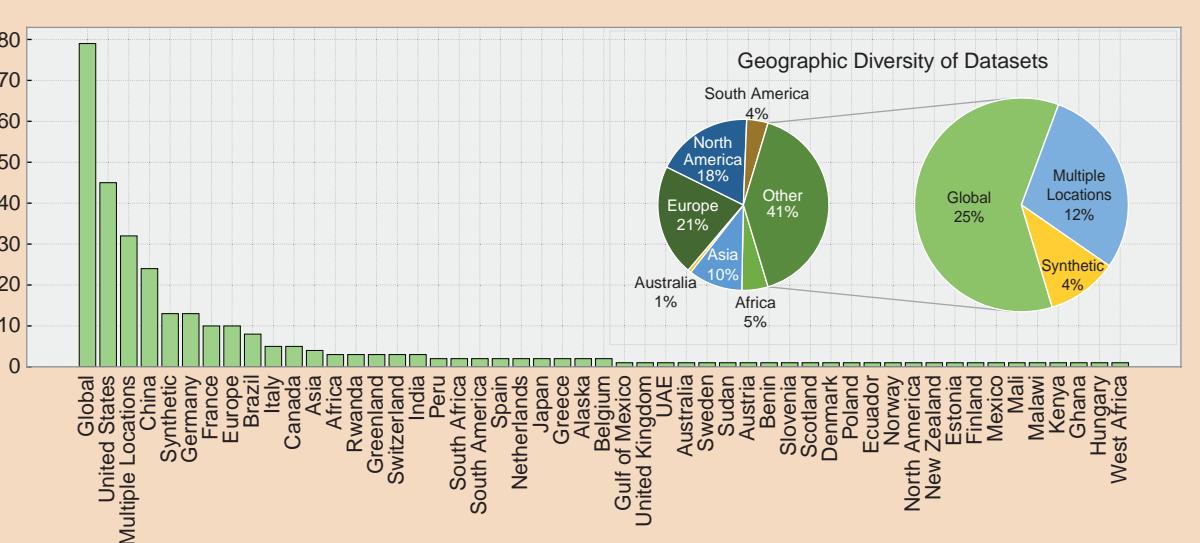


FIGURE 6. A geographic distribution of EO benchmark datasets (which provided clear location information). UAE: United Arab Emirates.

diverse datasets becomes apparent, in particular, covering underdeveloped countries as well.

SOURCE OF THE REFERENCE DATA

Another important aspect is the source of the reference data. Although most of the scientific papers that introduce a new benchmark dataset are very detailed regarding source and properties of the provided image data, they often contain only sparse information about the exact process of how the reference data were obtained. However, knowledge about the source of the reference data as well as about measures for quality assessment and quality control are essential to judge the potential of the benchmark and how to interpret obtained results. For example, achieving high “accuracy” on a dataset that contains a high number of annotation errors in its reference data only means that the model is reproducing the same bias as the method used to annotate the data but is not actually producing accurate results. Furthermore, information about the source of the reference data is not only scarce but also very heterogeneous. Examples include manual annotation for semantic segmentation [e.g., OpenEarthMap (Table 2; 253)] or object detection [e.g., BIRDSAI (Table 2; 127)], DroneCrowd (Table 2; 138), the use of keywords in image search engines for classification [e.g., AIDER (Table 2; 34)], leveraging existing resources [e.g., BigEarthNet (Table 2; 40)] uses the CORINE land use land cover (LULC) map, MapInWild (Table 2; 264) uses the World Database of Protected Areas, AI-TOD (Table 2; 162) uses other existing datasets, ship detection datasets usually employ AIS data, OpenSentinelMap (Table 2; 282) uses OSM, automatic reference methods (e.g., HSI Oil Spill

(Table 2; 211) and Sen12MS-CR-TS (Table 2; 98), or a mixture of these [e.g., RSOC (Table 2; 117)] uses existing datasets (DOTA, Table 2; 109) as well as manual annotations for new images.

The quality of class definitions varies among different datasets, often depending on whether a dataset is designed for a specific real-world application (in which class definitions are driven by the application context), or whether it is just an example dataset used to train and evaluate a ML-based approach (in which case, class definitions are more arbitrary). If human interaction is involved (e.g., via manual annotations), annotation protocols or precise class definitions are often not shared (if they even exist). With the evolution of datasets, the quality of the meta-information about the reference data needs to increase, together with the quality and quantity of the image data.

APPLICATIONS AND TASKS

Finally, we analyze a last aspect concerning the distribution of some characteristics of the datasets, namely, the number of classes and images. In Figure 7(a), we show the evolution of the number of classes per publication year. In Figure 7(b), we plot the number of classes against the number of images in the dataset. In both plots, the radius indicates the size of the image (width or height), while the color indicates the type of task. We can see that there is no clear trend or correlation between the year of publication and the number of classes in a dataset. Instead, we find that more recently published datasets increase the variety in the number of classes, again highlighting an increased interest in using benchmarking datasets in a wider range of applications.

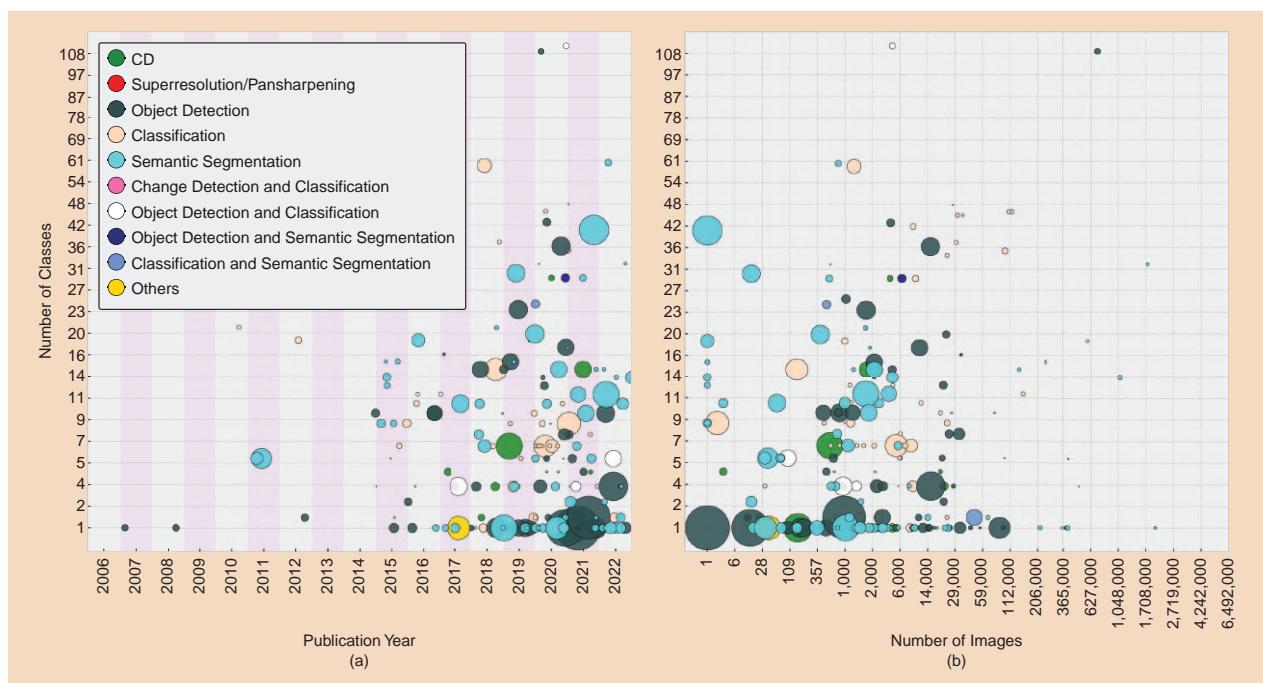


FIGURE 7. The number of classes provided by the reference data of a given dataset not only varies for different tasks (e.g., object detection is dominated by datasets with only a single class) but also with (a) publication year and (b) the number of images of a dataset.

Furthermore, although there is no clear correlation, we can confirm that increasing the number of classes reduces the size of the images and increases the number of images. More populated datasets, in terms of number of images, also have smaller image sizes and typically consider a smaller number of classes. Conversely, larger images are found in less populated datasets.

This overview is a snapshot of the currently published benchmarking datasets. Although the list of datasets continues to grow, we believe that the observed trends will not change much over the next few years. Instead, we expect long-term developments that will lead to two divergent aspects of datasets: specificity and generality. This will be further discussed in the “Open Challenges and Future Directions” section.

EXEMPLARY DATASETS

This section provides short descriptions of a selection of EO datasets for various tasks: semantic segmentation, classification, object detection, CD, and superresolution/pansharpening. Given the vast number of publicly available EO datasets, it is only possible to present some of them in this article. Thus, this selection cannot be comprehensive and certainly follows a subjective view influenced by the experience of the authors. However, the selected datasets are representative for their respective tasks and were selected based on their observed relevance: they either are the largest in terms of size (see the “Evolution of EO-Oriented ML Datasets” section for a definition), the first to be introduced for a given task, or the most popular dataset for the chosen application. The popularity was determined based on the number of citations the original paper introducing the dataset received.

The popularity of a dataset is influenced by multiple factors. One is certainly the size of a dataset, i.e., larger datasets are often preferred; however, there are exceptions. For

instance, Functional Map of the World (FMoW) (Table 2; 76), introduced in 2018 [14], is the largest dataset for remote sensing scene classification in terms of the number of images (1 million) but has yet to gain a high level of popularity [with 200 citations and 128 GitHub stars, compared to EuroSAT (Table 2; 52) with 796 citations and 276 GitHub stars, or AID (Table 2; 69) with 1,000 citations, which were all published in 2018]. Several other factors affect the popularity of a dataset, too, such as ease of access to the hosting servers

[Google Drive, IEEE DataPort, Amazon Web Services (AWS), university/personal servers, and so on], accompanying documentation and presentation (standard metadata format, detailed description, availability of template code and support, suitable search engine optimization, and so on), or ease of downloading the data (temporary or permanent links, bandwidth of hosting server, sign-in/up requirements, and so forth). Finally, an already-established dataset is more likely to be used in new studies to enable a comparison to related prior works, even if newer (and potentially better) datasets might exist. Along with brief descriptions, this section provides insights into the different dataset characteristics.

MORE POPULATED DATASETS, IN TERMS OF NUMBER OF IMAGES, ALSO HAVE SMALLER IMAGE SIZES AND TYPICALLY CONSIDER A SMALLER NUMBER OF CLASSES.

SEMANTIC SEGMENTATION

Semantic segmentation refers to assigning a class label to each pixel in the image. Partitioning the image into semantically meaningful parts creates a more detailed map of the input image and provides a foundation for further analysis such as LULC classification, building footprint extraction, landslide mapping, and so on [15]. Figure 8 shows the

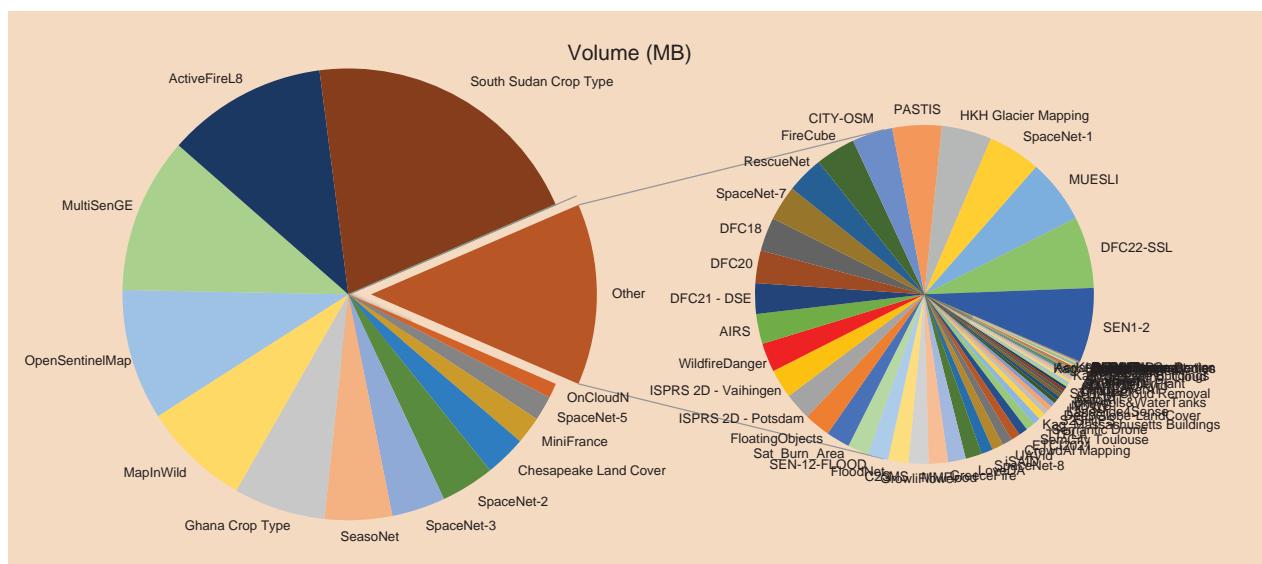


FIGURE 8. Relative volume distribution among datasets addressing semantic segmentation.

relative volume of corresponding benchmark datasets and illustrates a wide spread of dataset sizes. Here, we present two examples: SEN12MS (focusing on medium-resolution satellite data and weak annotations) and the ISPRS Vaihingen dataset (focusing on very high-resolution airborne data with high-quality annotations).

1) SEN12MS [49] is among the most popular and largest datasets (Table 2; 312) in terms of volume of data as shown in Figure 1. A total of 541,986 image patches of size 256×256 pixels with high spectral information content is present in this dataset [16]. It contains dual-polarimetric SAR image patches from *Sentinel-1*, multi-spectral image patches from *Sentinel-2*, and Moderate-Resolution Imaging Spectroradiometer (MODIS) land cover maps (Figure 9). The image patches are acquired

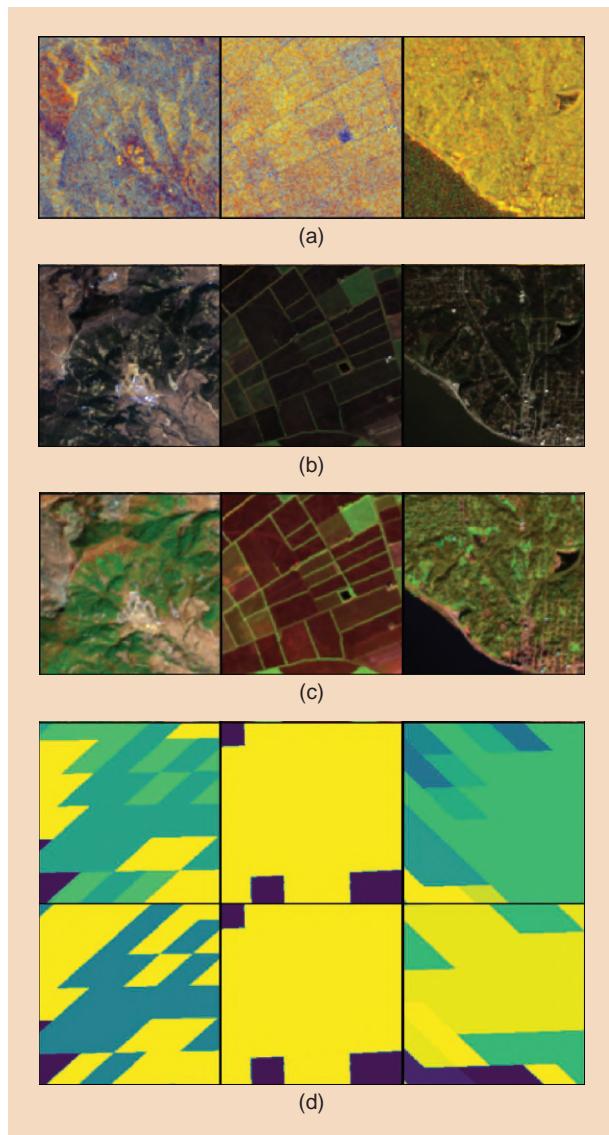


FIGURE 9. Several patch triplet examples from the SEN12MS dataset. (a) False-color *Sentinel-1* SAR (R: VV; G: VH; B: VV/VH), (b) *Sentinel-2* red, green, blue, (c) *Sentinel-2* short-wave infrared, (d) IGBP land cover, and LCCS land cover.

at random locations around the globe and cover the four meteorological seasons of the Northern Hemisphere. The patches were further processed to ensure that the images did not contain any clouds, shadows, and artifacts. In addition to the images and the land cover maps, the results of two baseline convolutional neural network classifiers (ResNet50 and DenseNet121) are also discussed to demonstrate the usefulness of the dataset for land cover applications [17].

2) ISPRS Vaihingen [18] is the earliest semantic segmentation dataset used for identifying land cover classes in aerial images [18], as shown in Figure 1 (Table 2; 220). This dataset contains a subset of images acquired by aerial cameras during the test of digital aerial cameras carried out by the German Association of Photogrammetry and Remote Sensing [19]. The dataset was prepared as part of a 2D semantic labeling contest organized by the International Society for Photogrammetry and Remote Sensing (ISPRS). The dataset contains images acquired over the city of Vaihingen in Germany. In total, orthorectified images of varying sizes and a digital surface model (DSM) are provided for 33 patches covering Vaihingen (Figure 10). The ground-sampling distance for the images and DSM is 9 cm. The three channels of the orthorectified images contain infrared, red, and green bands as acquired by the camera. The images are manually labeled for six common land cover classes: impervious surfaces, buildings, low vegetation, tree, car, and clutter/background.

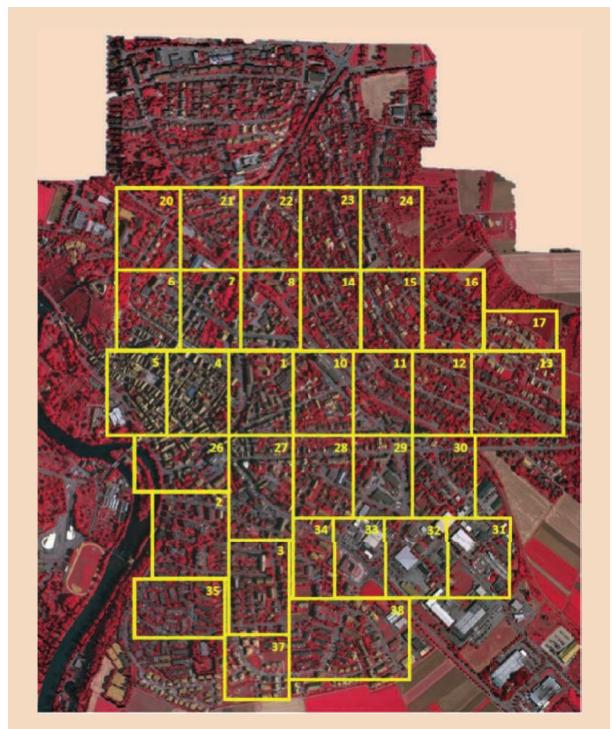


FIGURE 10. An illustration of 33 patches from the ISPRS Vaihingen dataset. (Source: [18].)

SCENE CLASSIFICATION

Remote sensing scene classification is closely related to semantic segmentation and has been used in various application domains such as urban planning [20], environment monitoring [21], LULC estimation [22], and so forth. The main difference is that instead of pixelwise classification and resolution-preserving maps as output, in scene classification, only one or more global labels are predicted for a given input image, aiming at a more comprehensive, generalized, and context-aware understanding of the underlying scene. Similar to image classification, which has been the driving force behind the early days of DL in computer vision, research on remote sensing scene classification has led to the creation of more diverse and large-scale high-resolution image datasets. Figure 11 shows the relative volume distribution of remote sensing scene classification datasets. This section covers EuroSAT as one of the earliest and FMoW as one of the largest datasets (representing roughly 50% of the available data in this task category) as well as BigEarthNet-MM, which introduces an additional interesting aspect by providing multiple labels for each image.

- ▶ EuroSAT [50] is one of the earliest large-scale datasets tailored for training deep neural networks for LULC classification tasks, as depicted in Figure 1 (Table 2; 52). The dataset includes 10 land cover classes (Figure 12), each containing 2,000–3,000 images, for 27,000 annotated and georeferenced images with 13 different spectral bands of 64×64 pixels. It contains multispectral images with a single label acquired from *Sentinel-2* satellite images of cities in 34 European countries [23], [24]. This dataset has been widely used for classification tasks, however, it may be used in a variety of real-world EO applications, such as detecting changes in land use and land cover as well as improving geographical maps as *Sentinel-2* images are freely available [23], [25], [26].

▶ BigEarthNet-MM [51] (Table 2; 40) is a benchmark archive that introduced an alternative nomenclature for images as compared to the traditional CORINE Land Cover (CLC) map for multilabel image classification and image-retrieval tasks. The CLC level-3 nomenclature is arranged into 19 classes for semantically coherent and homogenous land cover classes. This archive is created using *Sentinel-1* SAR and *Sentinel-2* multispectral satellite images acquired between June 2017 and May 2018 over 10 European countries [27]. The first version of BigEarthNet included only *Sentinel-2* image patches. It has been augmented with *Sentinel-1* image patches to form a multimodal benchmark archive (also called BigEarthNet-MM). The archive comprises 590,326 image patches with 12 spectral and two polarimetric bands. As shown in Figure 13, each image patch is annotated with several land cover classes (multilabels). One of the key features of the BigEarthNet dataset is its large number of classes and images, which makes it suitable for training DL models. However, because it is a multilabel dataset, the complexity of the problem is significantly increased compared to single-label datasets [28].

- ▶ FMoW [52] (Table 2; 76) is the largest dataset (Figure 1) for remote sensing scene classification in terms of the number of images [14]. The FMoW dataset is composed of approximately 1 million satellite images, along with metadata and multiple temporal views as well as subsets of the dataset that are classified into 62 classes, which are used for training, evaluation, and testing. Each image in the dataset comes with one or multiple bounding boxes indicating the regions of interest. Some of these regions may be present in multiple images acquired at different times, adding a temporal dimension to the dataset. The FMoW dataset is available to the public in two image formats: the FMoW-Full and the FMoW

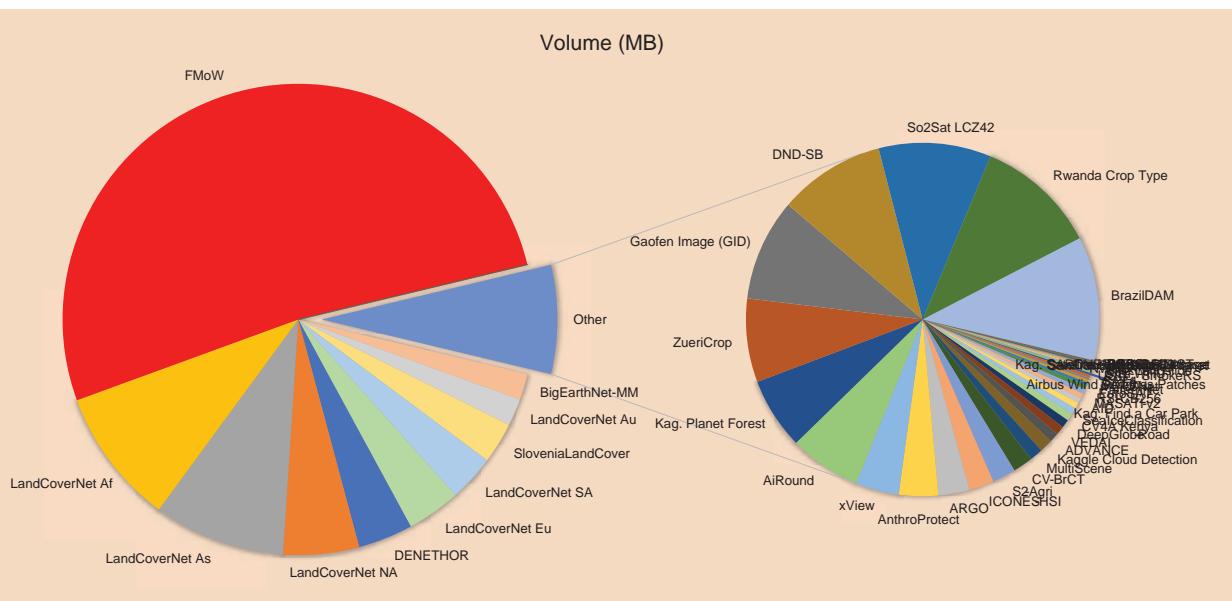


FIGURE 11. Relative volume distribution among datasets addressing scene classification.



FIGURE 12. Sample image patches of all 10 classes covered in the EuroSAT dataset.

red, green, blue (RGB). The FMoW-Full is in TIFF format and contains four- and eight-band multispectral imagery with a high spatial resolution resulting in 3.5 TB of data, while the FMoW-RGB has a much smaller size of

200 GB and includes all multispectral data converted to RGB in JPEG format. Examples of the classes in the dataset include flooded roads, military facilities, airstrips, oil and gas facilities, surface mines, tunnel openings, shipyards, ponds, and towers (see Figure 14 for examples). The FMoW dataset has a number of important characteristics, such as global diversity, multiple images per scene captured at different times, multispectral imagery, and metadata linked to each image.

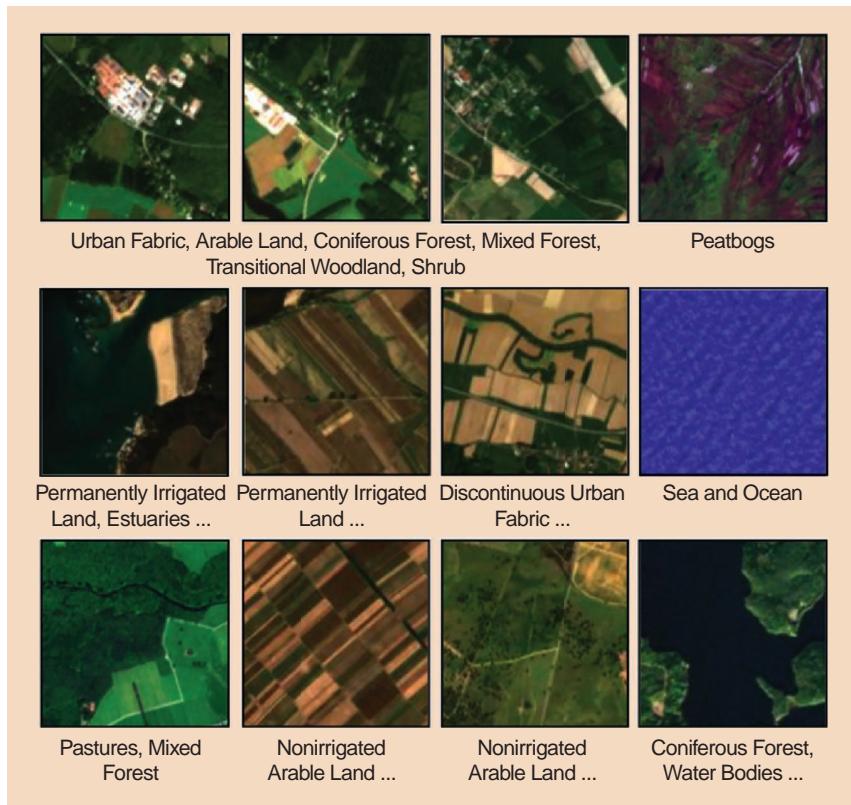


FIGURE 13. Sample image patches of several classes in the BigEarthNet-MM dataset with multiple labels being assigned to each image.

OBJECT DETECTION

The aim of object detection is to locate and identify the presence of one or more objects within an image, including objects with clear boundaries, such as vehicles, ships, and buildings, as well as those with more complex or irregular boundaries, for example, LULC parcels [29]. As seen in Figure 2, object detection is one of the most widely studied research tasks. Figure 15 shows the relative volume of the corresponding datasets. The xView3-SAR is by far the largest one. On the other hand, the family of DOTA datasets has a pioneering role in this field, placing them among the most popular datasets for object recognition tasks from remote sensing images.

The xView3-SAR [53] Ship Detection dataset is the largest labeled dataset (Table 2; 191), as shown in Figure 1, for training ML models to detect and classify ships in SAR images. The dataset includes nearly 1,000 analysis-ready

SAR images from the *Sentinel-1* mission, which are annotated using a combination of automated and manual analysis (see Figure 16 for image samples). The images are accompanied by co-located bathymetry and wind



FIGURE 14. Sample image patches for several classes from the FMoW dataset.

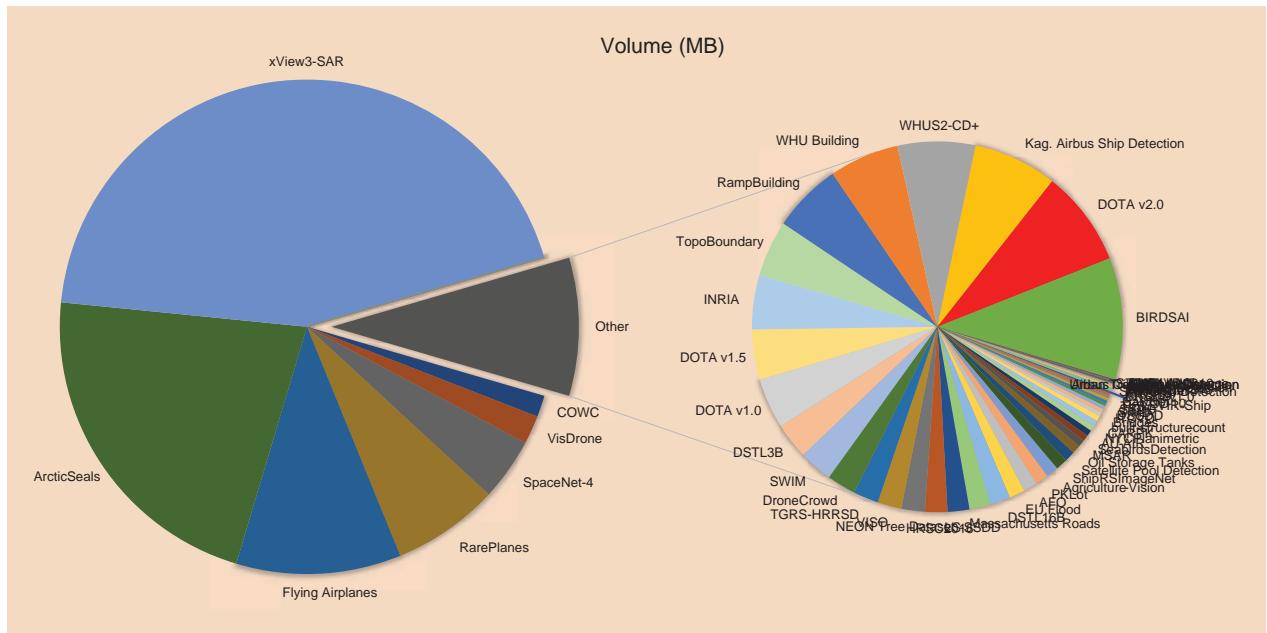


FIGURE 15. Relative volume distribution among the datasets addressing object detection.

state rasters, and the dataset is intended to support the development and evaluation of ML approaches for detecting “dark vessels” not visible to conventional monitoring systems.

- DOTA [54] is one of the most popular and largest object detection datasets (Table 2; 109 and 114) in terms of labeled object instances. It includes 2,806 images acquired from Google Earth (GE) and the China Center for Resources Satellite Data and Application [31], [32], [33].

The DOTA dataset is available in three different versions: DOTA-v1.0, DOTA-v1.5, and DOTA-v2.0. The image size in the initial version ranges from 800×800 pixels to $4,000 \times 4,000$ pixels, with 188,282 object instances with various scales and angular orientations and a total of 15 object categories. DOTA-v1.5 adds various small objects (fewer than 10 pixels) and a new container crane category with 402,089 instances, whereas DOTA-v2.0 adds two categories, airport and helipad, with 11,268 images and 1,793,658 instances, respectively. Some image samples of the DOTA dataset are presented in Figure 17.

CD

CD in remote sensing aims to identify temporal changes by analyzing multitemporal satellite images of the same location. CD is a popular task in EO as it fosters monitoring environmental changes through artificial or natural phenomena. Figure 2 shows that the number of dedicated CD datasets is small compared to other applications. Figure 18 shows that the available data are dominated by the DFC20 dataset (Track MSD), which focuses on semantic CD, followed by xView2, which tackles building damage assessment in the context of natural disasters. We chose LEVIR-CD as a recent dataset example and the Onera Satellite Change Detection dataset as one of the first large-scale datasets.

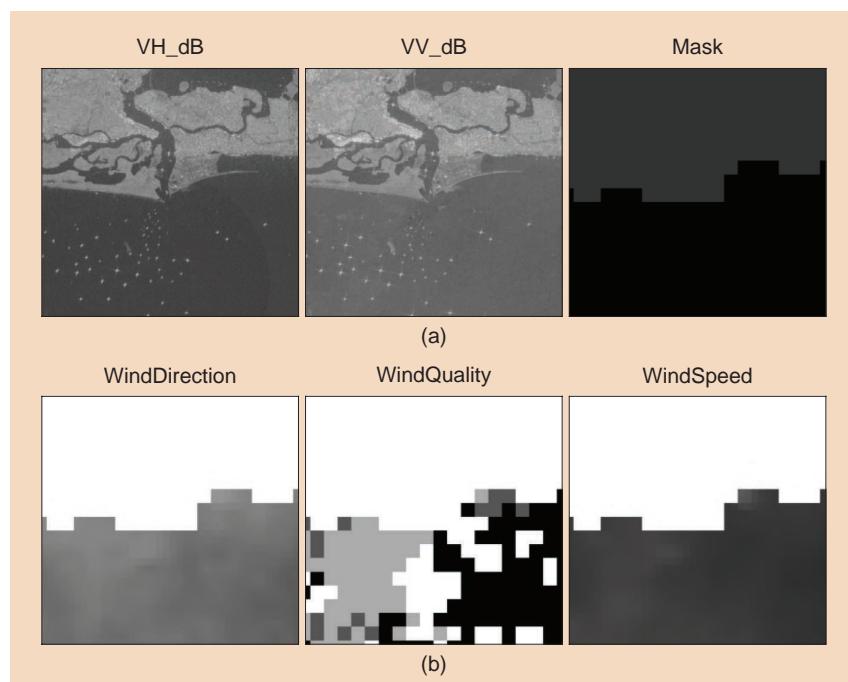


FIGURE 16. (a) An example image stack of dual-polarimetric SAR images and a water mask and (b) several wind properties from the xView3-SAR dataset.

- LEVIR-CD [55] is one of the most recent and largest CD datasets, as seen in Figure 1 (Table 2; 15). It was mainly developed for the evaluation of DL algorithms on building-related changes, including building growth (the transition from soil/grass/hardened ground, building under construction to new build-up regions) and building decline [33]. The dataset comprises 637 optical image patch pairings extracted from GE with a resolution of $1,024 \times 1,024$ pixels acquired over a time span of five to 14 years. LEVIR-CD covers various structures, including villa residences, tall apartments, small garages, and large warehouses, from 20 distinct regions in multiple Texas cities. The fully annotated LEVIR-CD dataset comprises 31,333 distinct change-building instances, some of which are illustrated in Figure 19, generated from the bitemporal images by remote sensing image interpretation specialists.
- Onera Satellite Change Detection [56] is one of the first, larger CD datasets, as shown in Figure 1 (Table 2; 11), containing multispectral image pairs from *Sentinel-2*. This dataset includes 24 pairs of multispectral images acquired from *Sentinel-2* satellites between 2015 and 2018, focusing on urban changes such

as new buildings or roads [34]. The locations were selected from around the world, including Brazil, the United States, Europe, the Middle East, and Asia. The reference data for pixel-level change are provided for

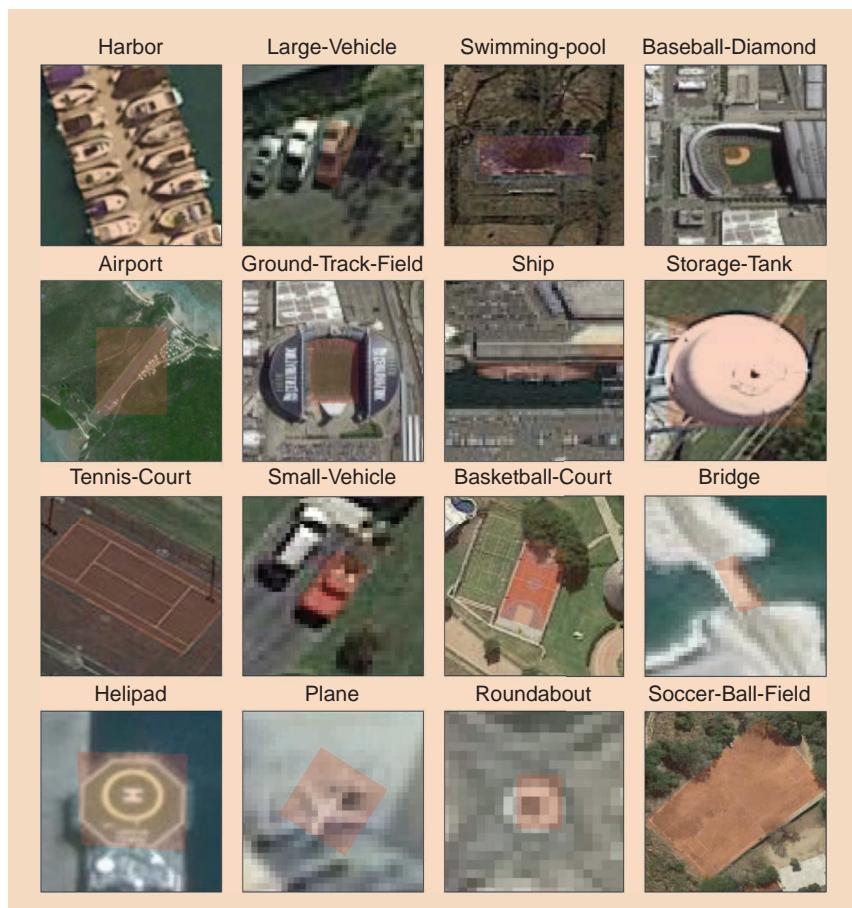


FIGURE 17. Several example image samples from the DOTA dataset.

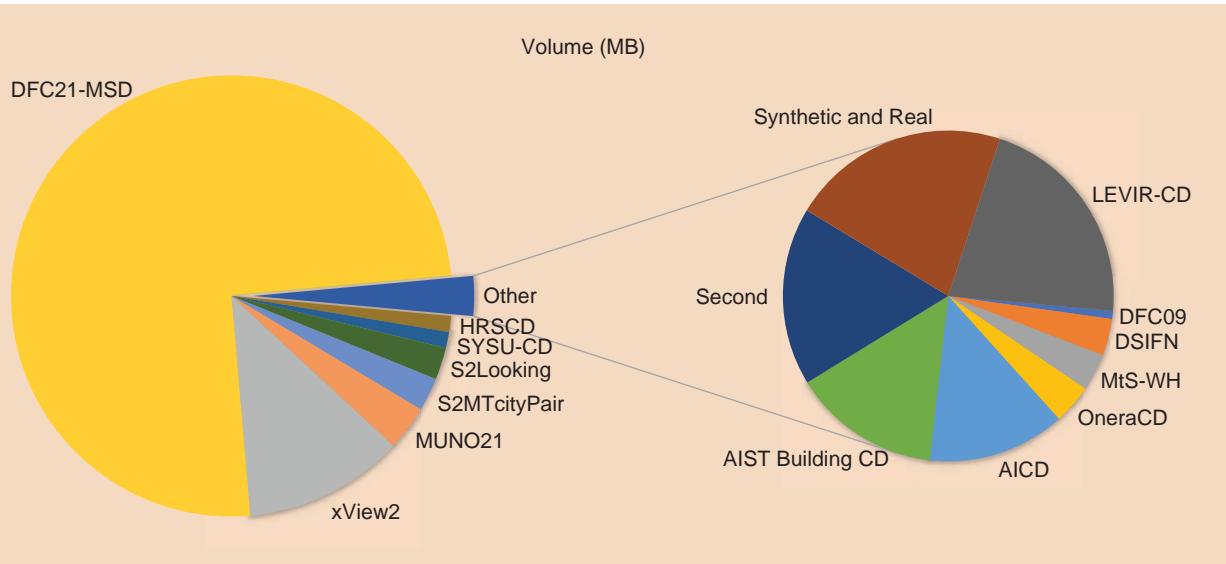


FIGURE 18. Relative volume distribution among the datasets addressing change detection.

all 14 training and 10 test image pairs. As illustrated in Figure 20, the annotated changes are primarily associated with urban changes, such as new buildings or roads.

SUPERRESOLUTION/PANSHARPENING

Pansharpening is one of the oldest data fusion approaches in remote sensing and aims to increase the spatial resolution of a multispectral image by combining it with



FIGURE 19. Examples of annotated samples from the LEVIR-CD dataset.

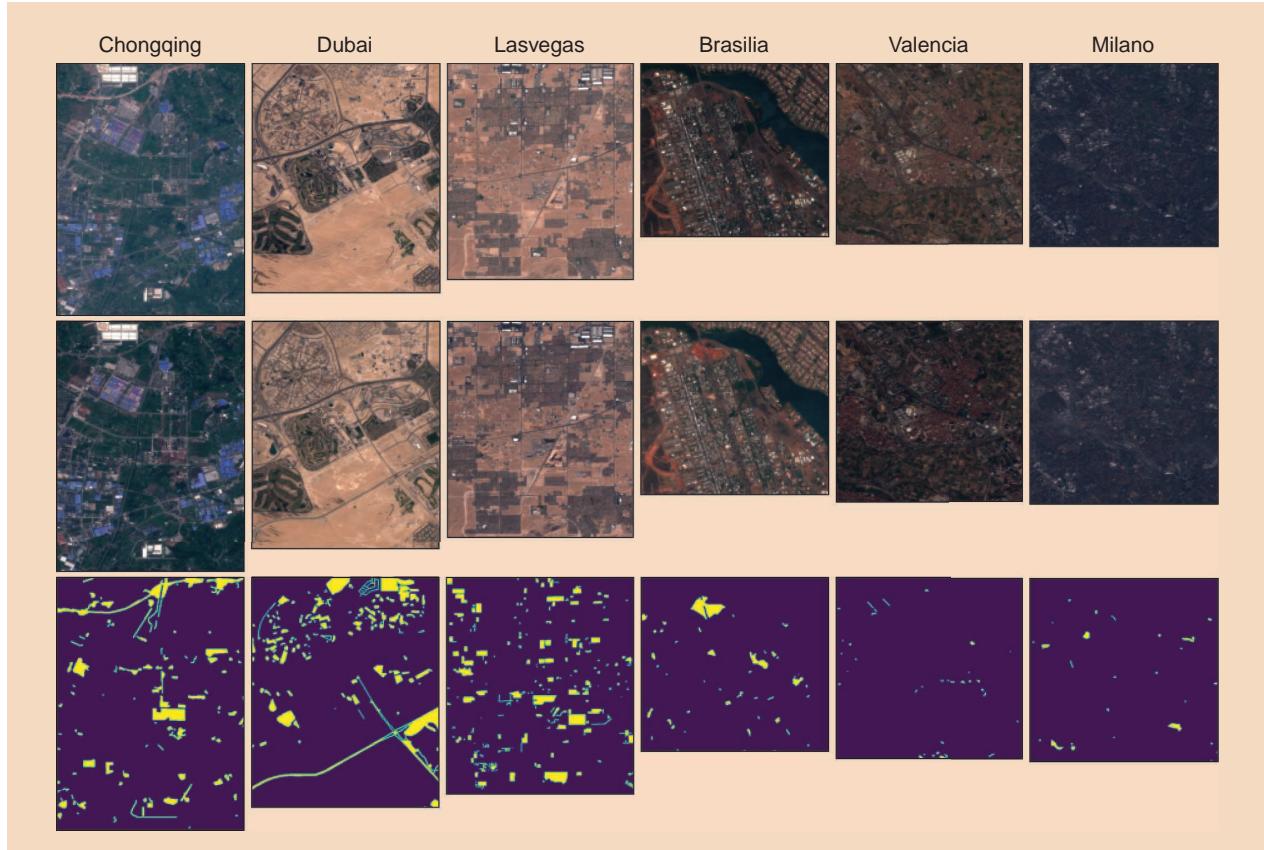


FIGURE 20. Examples of annotated samples from the Onera Satellite Change Detection dataset.

a panchromatic image. Due to the resolution difference between panchromatic and multispectral sensors, pansharpening is an exciting topic in remote sensing as it can provide a means to obtain higher-resolution data without better sensor technology. We select the Proba-V, PAirMax, and WorldStrat datasets as examples to showcase the peculiarities of datasets designed for that particular application.

1) Proba-V [57] is the earliest dataset available for superresolution [35] as Figure 1 shows (Table 2; 318). This dataset contains radiometrically and geometrically corrected Top-Of-Atmosphere reflectances in Plate Carré projection from the PROBA-V mission of the European Space Agency (Figure 21). The RED and near-infrared (NIR) spectral band data at 300 m and 100 m resolution are collected for 74 selected regions across the globe. Superresolution might be affected by the presence of pixels with clouds, cloud shadows, and ice/snow cover. Therefore, this dataset contains a quality map indicating pixels affected by clouds that should not be considered for superresolution. The dataset contains one 100 m resolution image and several 300 m resolution images of the same scene.

- 2) PAirMax [58] is a recently published (Table 2; 319 and 324) yet popular dataset for pansharpening. It contains 14 multispectral and panchromatic image pairs from six sensors on board satellite constellations of Maxar Technologies and Airbus [36]. As seen in Figure 22, most of the images are acquired over urban areas, highlighting several challenges for pansharpening (such as high-contrast features and adjacent regions with different spectral features). The work [36] also discusses the best practices to be followed in preparing high-quality, full-, and reduced-resolution images for pansharpening applications. In this dataset, the panchromatic band has 4×4 -times-more pixels than the multispectral bands. The multispectral bands are near the visible infrared region. The dataset also contains nine reduced-resolution test cases per Wald's protocol.
- 3) WorldStrat [59] is a recently introduced dataset for superresolution [37] and the largest in terms of volume

CD IS A POPULAR TASK IN EO AS IT FOSTERS MONITORING ENVIRONMENTAL CHANGES THROUGH ARTIFICIAL OR NATURAL PHENOMENA.

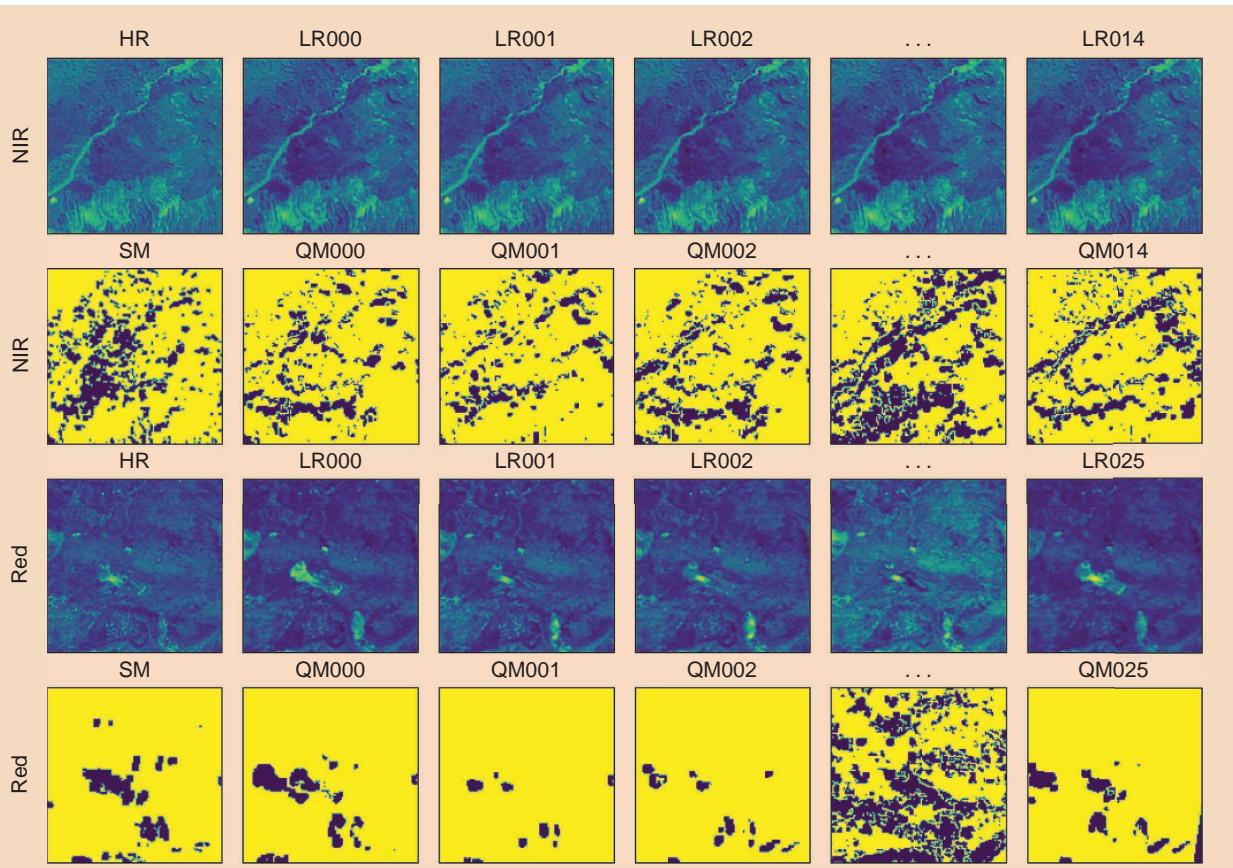


FIGURE 21. Sample images from the Proba-V dataset. Each sample consists of one high-resolution (HR) and several low-resolution (LR) images, each with a quality map (QM) showing which pixels are concealed (e.g., through clouds and so on).

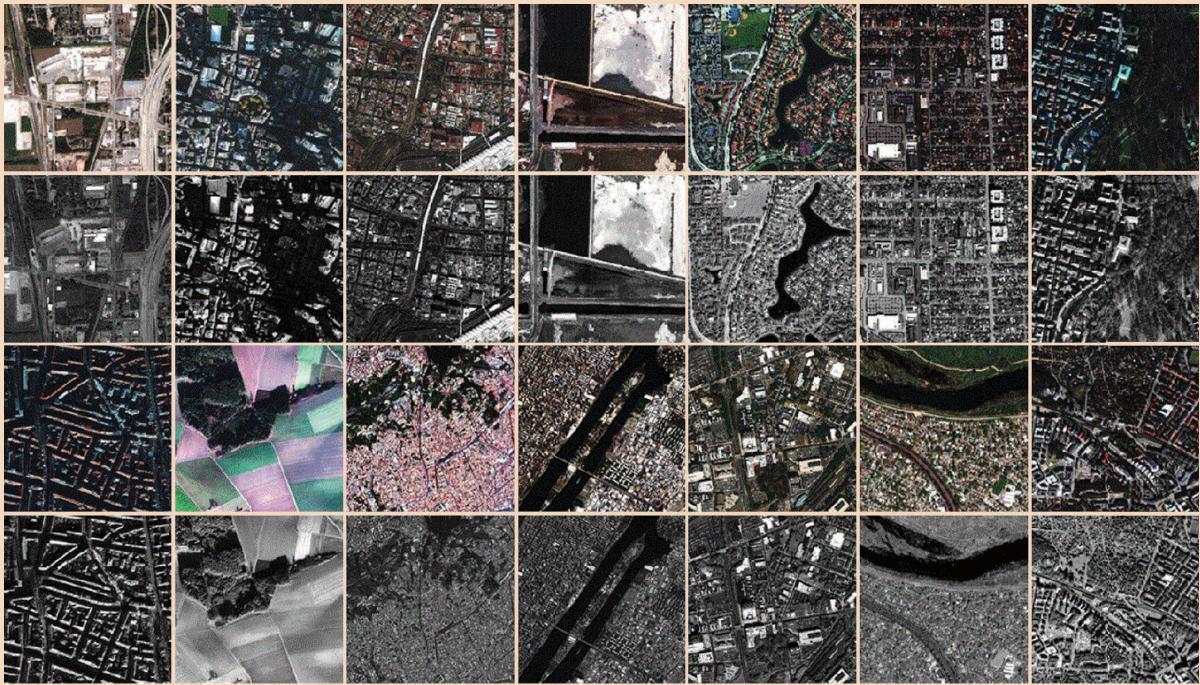


FIGURE 22. Sample images from PAirMax dataset. (Source: [58].)

(107 GB), as Figure 1 shows (Table 2; 325). This dataset contains high-resolution images from Airbus SPOT 6 and 7 along with 16 temporally matched low-resolution images from *Sentinel-2* satellites. The high-resolution images are over five spectral bands: the panchromatic band at a 1.5-m pixel resolution and RGB, and NIR bands at 6 m per pixel. The low-resolution ranges from 10 m per pixel to 60 m per pixel (Figure 23). In total, the dataset covers an area of approximately 10,000 km² and attempts to represent all types of land use across the world. Notably, the dataset contains nonsettlement and underrepresented locations such as illegal mining sites, settlements of persons at risk, and so on.

WORKING WITH REMOTE SENSING DATASETS

This section provides guidance on how to leverage available datasets to their full potential. Two of the main difficulties caused by information asymmetry (i.e., the information imbalance between the data providers and the data users) [5] have found suitable datasets and easy prototyping of ML approaches using such datasets. Here we discuss resources to gain an overview of existing datasets and download actual data, but we also provide examples of EO-oriented ML programming libraries.

DATA AVAILABILITY

Data availability covers two distinct aspects: on the one hand, access to the curated benchmark datasets, i.e., how

such datasets are made available to the public. This section provides several examples of the most common data sources. On the other hand, the actual noncurated measurements as acquired by the different sensors such as satellites, planes, and UAVs are often available too. Many data providers offer either their complete database for public access (e.g., the European Copernicus Open Access Hub [60]) or at least portions of their image acquisitions (e.g., via Open Data Programs such as those from Maxar [61] and Capella Space [62] or through scientific proposals

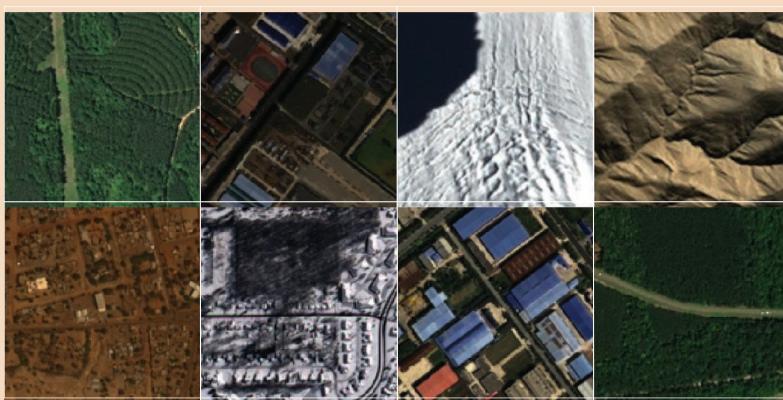


FIGURE 23. Sample images from the WorldStrat dataset. (Source: [37].)

as possible for *TerraSAR-X/TanDEM-X* data [63]). In principle, these data sources are highly valuable as they offer free access to petabytes of EO data, which can either be used to compile benchmark datasets by augmenting them with reference data for a specific task, or be leveraged for modern training paradigms such as self-supervised learning. An important aspect that is unfortunately sometimes ignored is the licensing of the corresponding data products: although direct usage (at least for scientific purposes) is usually free, redistribution of these data is often prohibited. Nevertheless, such image data find their way into public benchmark datasets, essentially causing copyright infringements. In addition to being in direct conflict with scientific integrity, such behavior is less likely to be tolerated in the future given the rising commercial value of EO imagery. Creators of future benchmark datasets should be fully aware of the license under which the leveraged EO data are accessible and how it limits their use in public datasets. In parallel, data providers should consider a licensing scheme that allows noncommercial redistribution for scientific/educational purposes to further facilitate the development and evaluation of modern ML techniques. The following is a listing of the most common data sources:

- ▶ IEEE DataPort [64] is a valuable and easily accessible data platform that enables users to store, search, access, and manage data. The platform is designed to accept all formats and sizes of datasets (up to 2 TB), and it provides both downloading capabilities and access to datasets in the cloud. Both individuals and institutions can indefinitely store and make datasets easily accessible to a broad set of researchers, engineers, and industry. In particular, most of the datasets used for previous IEEE GRSS Data Fusion Contests have been published on this platform (see “The Role of the IEEE Geoscience and Remote Sensing Society Data Fusion Contests.”) However, unless the dataset is associated with a competition or being submitted as open access, an IEEE account is required to access and download it.
- ▶ R radiant MLHub [65] enables users to access, store, register, and share high-quality open datasets for training ML models in EO. It’s designed to encourage widespread collaboration and the development of trustworthy applications. The available datasets in this platform cover research topics like crop type classification, flood detection, tropical storm estimation, and so on.
- ▶ Euro Data Cube (EDC) [66] provides a global archive of analysis-ready data (ARD) (Sentinel, Landsat, MODIS, and so forth), where users can search for and order data using the EDC Browser graphical interface (see Figure 24). It also enables users to store, analyze, and distribute user-contributed data content with simple application programming interfaces (APIs).
- ▶ NASA Earthdata [67] provides access to a wide range of remote sensing datasets, including satellite imagery,

atmospheric data, and land cover data, and offers a number of programming APIs. For example, the Application for Extracting and Exploring Analysis Ready Samples API enables users to perform data access and transformation processes easily and efficiently.

- ▶ Maxar Open Data Program [61] provides access to high-resolution remote sensing imagery collected since 2017, amounting to a total area of 1,834,152 km². It also features several programming APIs, such as the Maxar Catalog API and the ARD API. In addition, this program seeks to assist the humanitarian community by furnishing essential and useful information to support response efforts when crises occur.
- ▶ OpenAerialMap [68] is a community-driven platform that provides access to openly licensed satellite and UAV imagery from across the globe, accompanied by programming APIs for data access and processing. The platform currently hosts 12,614 images captured by 721 sensors, and all the images can be traced in OpenStreetMap [69].
- ▶ EarthNets [70] is an interactive data-sharing platform with more than 400 publicly available datasets in the geoscience and remote sensing field, which covers essential EO tasks like land use/cover classification, change/disaster monitoring, scene understanding, climate change, and weather forecasting [38]. Each benchmark dataset provides detailed meta information like spatial resolution and volume. Moreover, it also supports standard data loaders and codebases for different remote sensing tasks, which enables users to conduct a fair and consistent evaluation of DL methods on the benchmark datasets.
- ▶ EOD [71] is an interactive online platform for cataloging different types of datasets leveraging remote sensing imagery, which is developed by the IEEE GRSS IADF TC [6]. The key feature of EOD is to build a central catalog that provides an exhaustive list of available datasets with their basic information, which can be accessed and extended by the community and queried in a structured

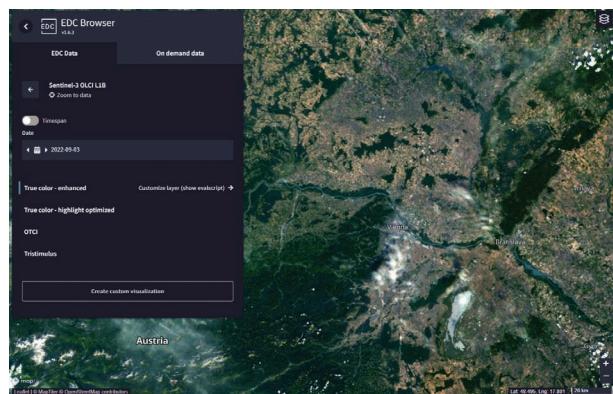


FIGURE 24. An illustration of the EDC Browser, where users can easily search for and order satellite data. (Source: [87].)

The Role of the IEEE Geoscience and Remote Sensing Society Data Fusion Contests

The IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest (DFC) has been organized as an annual challenge since 2006 by the IEEE GRSS Image Analysis and Data Fusion (IADF) Technical Committee (TC). The GRSS is an IEEE Society whose stated goal is to bring together researchers and practitioners to monitor and understand Earth's ecosystems and identify potential risks. The IADF is one of seven GRSS TCs aiming at technical contributions within the scope of geospatial image analysis [e.g., machine learning (ML), deep learning, image and signal processing, and big data] and data fusion (e.g., multisensor, multiscale, and multitemporal data integration). In general, the contest promotes the development of methods for extracting geospatial information from large-scale, multisensor, multimodal, and multitemporal data. It focuses on challenging problem settings and establishes novel benchmark datasets for open problems in remote sensing image analysis (see Table S1).

Historically, the DFC developed from less formal activities related to the distribution of datasets between 2002 and 2006 by means of a collaboration between the GRSS and the International Association of Pattern Recognition. In 2006, the first DFC was organized by what was then called the *Data Fusion TC of the GRSS*. It addressed the fusion of multispectral and panchromatic images, i.e., pansharpening, and provided one of the first public benchmark datasets for ML in remote sensing.

Since 2006, various sensors have played a role in the DFC, including optical (SPOT [S1], *DEIMOS-2* [S2], *WorldView-3* [S3], aerial [S4], [S5], [S6], [S7], [S8], [S9]), multispectral (*Landsat* [S10], [S11], [S12], [S13], *Worldview-2* [S14], [S15], *Quickbird* [S15], *Sentinel-2* [S11], [S13], [S16], and aerial [S12]) as well as hyperspectral images [S7], [S17], [S18], SAR (ERS [S1], [S10], *TerraSAR-X* [S15], *Sentinel-1* [S13], [S16] *Gaofen* [S9]) and lidar [S5], [S6], [S7], [S14], [S17], [S18] data, but also less common sources of Earth observation (EO) data such as thermal images [S4], digital elevation models [S8], video from space (ISS [S2]),

nighttime images (Visible Infrared Imaging Radiometer Suite [S13]), and OpenStreetMap [S11]. Another meaningful change happened in 2017 when the DFC moved away from providing data over a single region only but instead allowed the training of models over five cities worldwide (Berlin, Hong Kong, Paris, Rome, and São Paulo) and testing on four other cities (Amsterdam, Chicago, Madrid, and Xi'an). This enabled the creation of models that generalize to new and unseen geographic areas instead of overfitting to a single location. This commendable trend was continued in 2020 by using SEN12MS (Table 2; 87) [S20] as training data [S16], providing data for the whole state of Maryland, USA, in 2021 (Table 2; 1 and 291) [S12], in total, 19 different conurbations in France in 2022 (Table 2; 241) [S8], and data from 17 cities across six continents in 2023 [S9].

The tasks addressed by the DFC over the years have been dominated by semantic mapping [S3], [S4], [S5], [S6], [S7], [S9], [S10], [S11], [S13], [S17], [S18], [S19] but also include pansharpening (Table 2; 320) [S21], change detection (in the context of floods) (Table 2; 10) [S1] and 3D reconstruction (Table 2; 193) [S3], [S9], modern challenges such as weakly (Table 2; 1/287/291) [S12], [S16] and semisupervised learning (Table 2; 241) [S8] as well as open task contests [S2], [S4], [S14], [S15], which allowed participants to freely explore the potential of new and uncommon EO data.

In 2006, seven teams from four different countries participated in the first DFC [S21] despite public contests being a new concept within the EO community. Being organized by an internationally well-connected Society, providing exciting challenges, and establishing new benchmarks led to a quick rise in popularity. From seven teams in four countries in 2006, participation jumped quickly to 21 teams in 2008, 42 teams in 2014, and reached its peak with 148 teams (distributed over four different tracks, however) in 2019. The peak of the popularity was around 2012, when the DFC attracted

TABLE S1. AN OVERVIEW OF THE IEEE GRSS DATA FUSION CONTESTS FROM 2006 TO 2023.

YEAR	DATA	GOAL	REFERENCE
2023	Very high-resolution (VHR) optical and SAR satellite images	Classification, height estimation	[S9]
2022	VHR aerial optical images, DEM	Semisupervised learning	[S8]
2021	Multitemporal multispectral (aerial and <i>Landsat-8</i>) imagery	Weakly supervised learning	[S12]
	Multispectral (<i>Landsat-8</i> , <i>Sentinel-2</i>), SAR (<i>Sentinel-1</i>), nighttime (Visible Infrared Imaging Radiometer Suite) images	Semantic segmentation	[S13]
2020	Multispectral (<i>Sentinel-2</i>) and SAR (<i>Sentinel-1</i>) imagery	Weakly supervised learning	[S16]
2019	Optical (<i>Worldview-3</i>) images and lidar data	Semantic 3D reconstruction	[S3], [S19]
2018	Multispectral lidar data, VHR optical and hyperspectral imagery	Semantic segmentation	[S7]
2017	Multispectral (<i>Landsat</i> , <i>Sentinel-2</i>) images and OpenStreetMap	Semantic segmentation	[S11]
2016	Very high temporal resolution imagery (<i>DEIMOS-2</i>) and video from space (ISS)	Open for creative ideas	[S2]
2015	Extremely high-resolution lidar and optical data	Semantic segmentation	[S5], [S6]
2014	Coarse resolution thermal/hyperspectral data and VHR color imagery	Semantic segmentation	[S4]
2013	Hyperspectral imagery and a lidar-derived digital surface model	Semantic segmentation	[S18]
2012	VHR optical (<i>QuickBird</i> and <i>WorldView-2</i>), SAR (<i>TerraSAR-X</i>), and lidar data	Open for creative ideas	[S15]
2011	Multiangular optical images (<i>WorldView-2</i>)	Open for creative ideas	[S14]
2009-2010	Multitemporal optical (SPOT) and SAR (ERS) images	Change detection	[S1]
2008	VHR hyperspectral imagery	Semantic segmentation	[S17]
2007	Low-resolution SAR (ERS) and optical (<i>Landsat</i>) data	Semantic segmentation	[S10]
2006	Multispectral and panchromatic images	Pansharpening	[S21]

ERS: European Remote Sensing.

more than 1,000 registrations for the contest from nearly 80 different countries. Influenced by different factors, including the overwhelming success of datasets (and connected contests) in the Computer Vision community, an increasing number of EO sensors with easier access to their data, as well as improved options for data hosting, the number of available benchmark datasets (that were not always but often connected to a contest) increased dramatically around 2015 (see also Figure 1). Interestingly, the growing influence of Computer Vision on remote sensing, in particular due to deep learning, is also reflected by renaming the Data Fusion Technical Committee to the Image Analysis and Data Fusion (IADF) Technical Committee in 2014. Since then, participation in the DFC has been more or less constant (with a few positive outliers such as 2019), with approximately 40 teams from roughly 20 countries. Another commendable fact is that at the beginning, participants of the DFC were dominated by well-established scientists with solid experience in their respective fields. Although this group still plays a vital role in more recent DFCs, a large number of participants (and winners!) are students. This shows that improved data availability helped lower the starting threshold to analyze various EO data by standardizing data formats, easing access to theoretical knowledge, and open sourcing software libraries and tools.

References

- [S1] N. Longbotham et al., "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012, doi: 10.1109/JSTARS.2011.2179638.
- [S2] L. Mou et al., "Multitemporal very high resolution from space: Outcome of the 2016 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3435–3447, Aug. 2017, doi: 10.1109/JSTARS.2017.2696823.
- [S3] S. Kunwar et al., "Large-scale semantic 3D reconstruction: Outcome of the 2019 IEEE GRSS data fusion contest - Part A," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 922–935, 2021, doi: 10.1109/JSTARS.2020.3032221.
- [S4] W. Liao et al., "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2984–2996, Jun. 2015, doi: 10.1109/JSTARS.2015.2420582.
- [S5] M. Campos-Taberner et al., "Processing of extremely high-resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest–part A: 2-D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5547–5559, Dec. 2016, doi: 10.1109/JSTARS.2016.2569162.
- [S6] A.-V. Vo et al., "Processing of extremely high resolution LiDAR and RGB data: Outcome of the 2015 IEEE GRSS data fusion contest–Part B: 3-D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5560–5575, Dec. 2016, doi: 10.1109/JSTARS.2016.2581843.
- [S7] Y. Xu et al., "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1709–1724, Jun. 2019, doi: 10.1109/JSTARS.2019.2911113.
- [S8] R. Hänsch et al., "The 2022 IEEE GRSS data fusion contest: Semisupervised learning [Technical Committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 334–337, Mar. 2022, doi: 10.1109/MGRS.2022.3144291.
- [S9] C. Persello et al., "2023 IEEE GRSS data fusion contest: Large-scale fine-grained building classification for semantic urban reconstruction [Technical Committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 1, pp. 94–97, Mar. 2023, doi: 10.1109/MGRS.2023.3240233.
- [S10] F. Pacifici, F. Del Frate, W. J. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Outcome of the 2007 GRSS data fusion contest," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 331–335, Jul. 2008, doi: 10.1109/LGRS.2008.915939.
- [S11] N. Yokoya et al., "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, May 2018, doi: 10.1109/JSTARS.2018.2799698.
- [S12] Z. Li et al., "The outcome of the 2021 IEEE GRSS data fusion contest–Track MSD: Multitemporal semantic change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1643–1655, Jan. 2022, doi: 10.1109/JSTARS.2022.3144318.
- [S13] Y. Ma et al., "The outcome of the 2021 IEEE GRSS data fusion contest - Track DSE: Detection of settlements without electricity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12,375–12,385, Nov. 2021, doi: 10.1109/JSTARS.2021.3130446.
- [S14] F. Pacifici and Q. Du, "Foreword to the special issue on optical multiangular data exploitation and outcome of the 2011 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 3–7, Feb. 2012, doi: 10.1109/JSTARS.2012.2186733.
- [S15] C. Berger et al., "Multi-modal and multi-temporal data fusion: Outcome of the 2012 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1324–1340, Jun. 2013, doi: 10.1109/JSTARS.2013.2245860.
- [S16] C. Robinson et al., "Global land-cover mapping with weak supervision: Outcome of the 2020 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3185–3199, Mar. 2021, doi: 10.1109/JSTARS.2021.3063849.
- [S17] G. Licciardi et al., "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009, doi: 10.1109/TGRS.2009.2029340.
- [S18] C. Debes et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014, doi: 10.1109/JSTARS.2014.2305441.
- [S19] Y. Lian et al., "Large-scale semantic 3-D reconstruction: Outcome of the 2019 IEEE GRSS data fusion contest–Part B," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1158–1170, 2021, doi: 10.1109/JSTARS.2020.3035274.
- [S20] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS - A curated data-set of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 153–160, Sep. 2019, doi: 10.5194/isprs-annals-IV-2-W7-153-2019.
- [S21] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007, doi: 10.1109/TGRS.2007.904923.

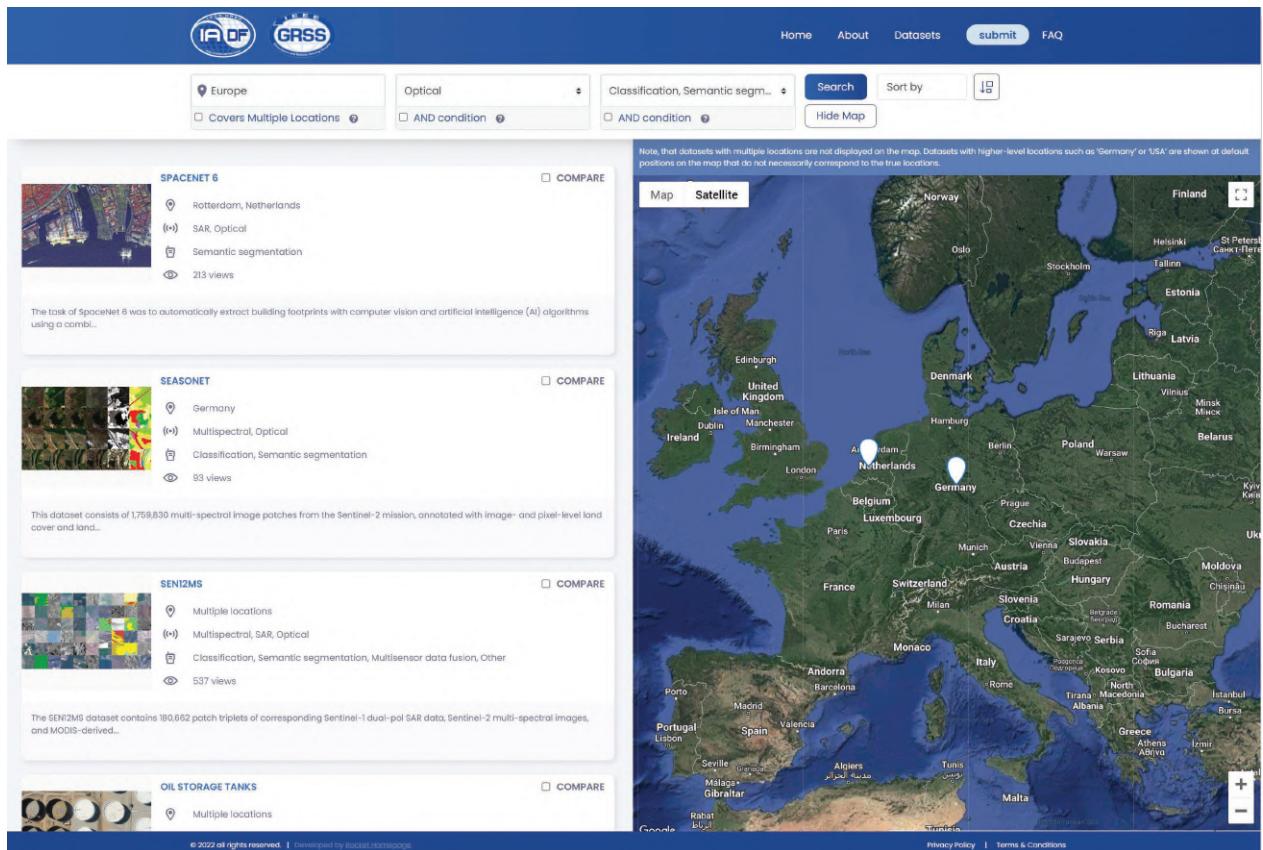


FIGURE 25. An illustration of the map view of the EOD data catalog.

and interactive manner (see Figure 25 for an example of the user interface). For more information, see “The IEEE Geoscience and Remote Sensing Society Earth Observation Database.”

EO-ORIENTED ML LIBRARIES

Most of the existing ML libraries are developed for classic computer vision tasks, where the input image is usually single-channel (grayscale) or with RGB bands. EO datasets are often of large volumes with highly diverse data types, different numbers of spectral bands, and spatial resolutions, as illustrated in Figure S1. The code base for processing such data samples is often highly complex and difficult to maintain. One approach to increase readability, reusability, and maintainability is to modularize the code and encapsulate different tasks by decoupling processing the dataset (data loaders, visualization, preprocessing, and so on) and applying ML models (training, prediction, model selection, evaluation, and so forth). A major challenge in training advanced ML models for EO tasks is the implementation of an easy-to-use, yet efficient data loader explicitly designed for geoscience data that loads and preprocesses a complex dataset and produces an iterable list of data samples in a customizable way.

This section introduces several well-known packages designed explicitly for geoscience and remote sensing

data. As PyTorch [39] and TensorFlow [40] (note that Keras is now a part of TensorFlow) are the most widely used DL frameworks, we focus mainly on the introduction for the existing libraries, using these two DL frameworks as the back end.

- 1) TorchGeo [72] is an open source PyTorch-based library, which provides datasets, samplers, transforms, and pretrained models specific to geospatial data [41]. The main goal of this library is to simplify the process of interacting with complex geospatial data and make it easier for researchers to train ML models for EO tasks. Figure 26 provides an example of sampling pixel-aligned patch data from heterogeneous geospatial data layers using the TorchGeo package. As different layers usually have different coordinate systems and spatial resolutions, patches sampled from these layers in the same area may not be pixel aligned. Therefore, in a practical application scenario, researchers need to conduct a series of pre-processing operations such as reprojecting and resampling of the geospatial data before training ML models, which is time consuming and laborious. To address this challenge, TorchGeo provides data loaders tailored for geospatial data, which support transparently loading data from heterogeneous geospatial data layers with relatively simple code, per the following example:

```

1 From torch.utils.data import DataLoader
2 from torchgeo.datasets import CDL, Landsat8, stack_samples
3 from torchgeo.samplers import RandomGeoSampler
4
5 # Loading Landsat8 and CDL layers
6 landsat8 = Landsat8(root="...")
7 cdl = CDL(root="...", download=True,
checksum=True)
8
9 # Take the intersection of Landsat8 and CDL
10 dataset = landsat8 & cdl
11
12 # Sample 10,000 256 x 256 image patches
13 sampler = RandomGeoSampler(dataset, size
= 256, length = 10000)
14
15 # Build a normal PyTorch DataLoader with
the sampler
16 dataloader = DataLoader(dataset,
batch_size = 128, sampler=sampler,
collate_fn=stack_samples)
17
18 for batch in dataloader:
19     image = batch["image"]
20     mask = batch["mask"]
21
22 # Train a model
23 ...

```

A more detailed introduction about the supported geospatial datasets in TorchGeo can be found in [41].

- 2) RasterVision [73] is an open source Python framework that aims to simplify the procedure for building DL-based computer vision models on satellite, aerial, and other types of geospatial data (including oblique drone imagery). It enables users to efficiently construct a DL pipeline, including training data preparation, model training, model evaluation, and model deployment, without any expert knowledge of ML. Specifically, RasterVision supports chip classification, object detection, and semantic segmentation with the PyTorch back end on both CPUs and GPUs, with built-in support for running in the cloud using AWS. The framework is extensible to new data sources, tasks (e.g., instance segmentation), back ends (e.g., Detectron2), and cloud providers. Figure 27 provides the pipeline of the RasterVision package. A more comprehensive tutorial for this package can be found in [74].
- 3) Keras Spatial [75] provides data samplers and tools designed to simplify the preprocessing of geospatial data for DL applications with the Keras back end. It provides a data generator that reads samples directly from raster layers without creating small patch files before run-

The IEEE Geoscience and Remote Sensing Society Earth Observation Database

The core of the Earth Observation Database (EOD) is its database of user-submitted datasets. It can be searched by a combination of various queries, including location as well as different combinations of sensors and tasks. Search results (as well as the whole database) can be viewed either in an illustrated list view or in an interactive map view, indicating those datasets that cover only a specific location with a marker (see Figure 25). Clicking on one of the markers in the map view will limit the list to datasets at this specific location. Clicking on a dataset in the list will open a new window, displaying detailed information about this dataset, which includes

- ▶ geographic location
- ▶ sensor modality
- ▶ task/application
- ▶ data size
- ▶ a Uniform Resource Locator to access the data
- ▶ number of views
- ▶ a graphical representation
- ▶ a brief description of the dataset.

A helpful function is the compare option, which allows a direct side-to-side comparison of this information from two or more datasets in a newly opened window. EOD is meant as a community-driven catalog maintained by the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion (IADF) Technical Committee, i.e., adding datasets is neither limited to IADF members nor to the creators of the dataset, but to anybody with an interest who increases the visibility and accessibility of a certain dataset can request to include it in EOD. The submission requests are reviewed by the IADF for completeness and correctness before the dataset is added to the database and visible to the public.

ning the model. It supports loading raster data from local files or cloud services. Necessary preprocessing, like reprojecting and resampling, is also conducted automatically. Keras Spatial supports sample augmentation

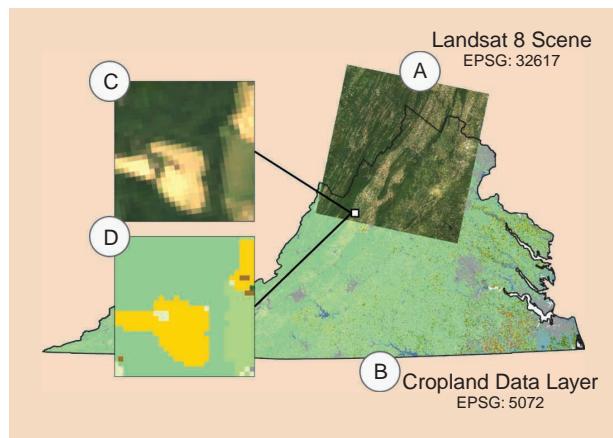


FIGURE 26. An illustration of sampling from heterogeneous geospatial data layers. (Source: [41].) With the TorchGeo package, users can focus directly on training ML models without manually preprocessing the heterogeneous geospatial data, including aligning data layers by reprojection and resampling.

using a user-defined callback system to improve the flexibility of data management. Here, a simple demo code using the `SpatialDataGenerator` class from Keras Spatial to prepare a training set for a DL model is given:

```

1  from keras_spatial import SpatialData-
2    Generator
3
4  # Loading labels from a local file
5  labels = SpatialDataGenerator()
6  labels.source = '/path/to/labels.tif'
7  # Sample 128 x 128 patches
8  labels.width, labels.height = 128, 128
9  # Set a geodataframe with 200 x 200
10 # in projection units of the original raster
11 df = labels.regular_grid(200,200)
12
13 # Loading images from a file on the cloud
14 samples = SpatialDataGenerator()
15 samples.source = 'https://server.com/
16 files/data.tif'
17 samples.width, samples.height = labels.
18 width, labels.height
19
20 # The training set generator
21 train_gen = zip(labels.flow_from_
22      _dataframe(df), patches.flow_from_
23      _dataframe(df))
24
25 # Train a model
26 model(train_gen)

```

GEOSPATIAL COMPUTING PLATFORMS

In addition to ML libraries, public geospatial computing platforms such as Google Earth Engine (GEE) for EO tasks in practical application scenarios offer a series of benefits, including the following:

- ▶ *Access to large-scale datasets*: geospatial computing platforms usually provide access to large and diverse geospatial datasets, such as satellite imagery, weather data, and terrain data.

These datasets may be time consuming and expensive to acquire on one's own but can be easily accessed through a geospatial computing platform using the cloud service.

▶ *Scalability*: geospatial computing platforms are designed to address large-scale geospatial data processing and analysis. They usually provide cloud-based computing resources that can be easily scaled up or down to meet researchers' needs. This makes it easier to perform complex geospatial analysis that would be difficult to do on local machines considering the limited computing resources.

▶ *Prebuilt tools and APIs*: geospatial computing platforms usually provide prebuilt tools and programming APIs for image processing, geocoding, and data visualization, making it much easier for researchers to work with geospatial data.

Several representative geospatial computing platforms are as follows:

▶ GEE [76] is a cloud-based platform designed for large-scale geospatial data analysis and processing. It provides a range of tools and APIs that allow users to analyze and visualize geospatial data, including raster and vector processing tools, ML algorithms, and geospatial modeling tools. In addition, GEE provides access to powerful computing resources, such as virtual machines and storage, to enable users to perform complex geospatial analyses. The GEE Data Catalog contains more than 40 years of historical imagery and scientific datasets for Earth science, which are updated and expanded daily. These datasets cover various topics such as climate, weather, surface temperature, terrain, and land cover. Notable datasets available on the GEE Data Catalog include Planet SkySat Public Ortho Imagery (collected for crisis response events) [42] and NAIP [77] (agricultural monitoring data in the United States).

▶ AWS [78] is a powerful platform for geospatial computing that offers a range of services for geospatial data storage, processing, and analysis. AWS hosts several representative geospatial datasets, including Digital Earth Africa [79] (*Landsat* and *Sentinel* products over Africa), National Oceanic and Atmospheric Administration Emergency Response Imagery [80] (lidar and hyperspectral data over the United States), and datasets from the SpaceNet

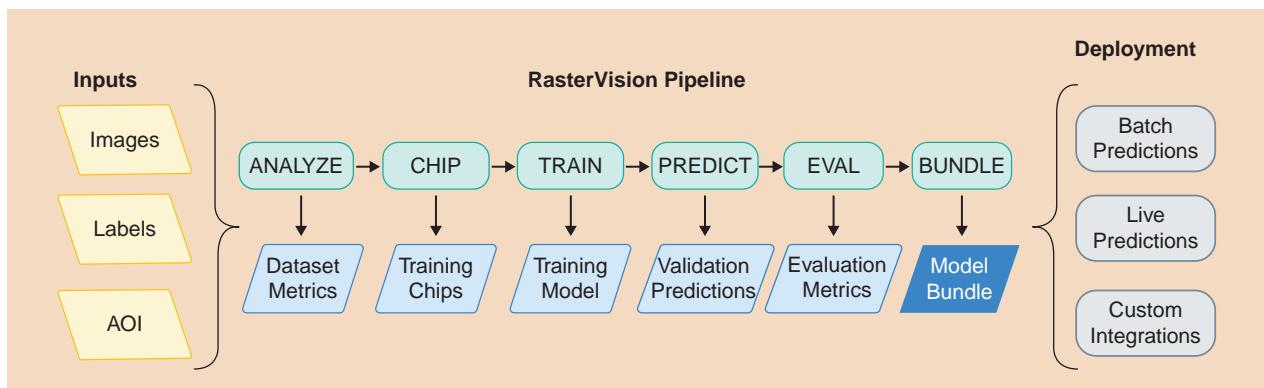


FIGURE 27. An illustration of the pipeline from the RasterVision package. AOI: area of interest. (Source: <https://github.com/azavea/raster-vision>.)

challenges (Table 2; 134, 264, 298, 299, 300, 304, 306, and 312) [43]. Sharing data on AWS allows anyone to analyze them and build services using a broad range of computing and data analytical products like Amazon EC2, which enables data users to spend more time on data analysis rather than data acquisition.

- Microsoft Planetary Computer [81] provides access to a wide range of EO data and powerful geospatial computing resources. The platform is specifically designed to support the development of innovative applications and solutions for addressing critical environmental challenges, such as climate change, biodiversity loss, and natural resource management. It offers cloud-based computing services that enable users to perform complex geospatial analyses efficiently. The Planetary Computer Data Catalog provides access to petabytes of environmental monitoring data in a consistent and analysis-ready format. Some of the representative datasets available on the Planetary Computer Data Catalog include the HREA dataset [82] (settlement-level measures of electricity access derived from satellite images) and the Microsoft Building Footprints dataset [83].
- Colab [84] provides a flexible geospatial computing platform that offers users access to free GPU and tensor processing units (TPUs) computing resources for analyzing geospatial data. It provides a free Jupyter notebook environment that allows users to write and run Python code for accessing and processing geospatial data from various platforms (e.g., the GEE Data Catalog). Colab notebooks can be easily shared and collaborated on with others, which is particularly useful for geospatial analysis projects that involve multiple team members.
- Kaggle [86] is an online platform widely used for data science and ML competitions. It provides a cloud-based computational environment that allows for reproducible and collaborative analysis in the field of geospatial computing. Kaggle also provides access to a variety of geospatial datasets, including satellite imagery, terrain data, and weather data, along with tools to analyze and visualize these datasets. Users can take advantage of the platform's free GPU and TPU computing resources to train ML models and undertake complex geospatial analysis tasks.

OPEN CHALLENGES AND FUTURE DIRECTIONS

The previous sections give an overview of existing benchmarking datasets, presenting their main features and describing their main characteristics, eventually providing a broad yet detailed picture of the current state of the art. This section discusses existing gaps, open challenges, and potential future trends.

WHERE ARE WE NOW?

The most prominent issue that new dataset releases aim to fix is a lack of diversity. Many of the earlier datasets contain

samples that are extremely limited in their spatial and temporal distribution, to the extreme of consisting of a single, small image only. Not only are such datasets prone to leading to biased evaluation protocols, where information from the training set leaks into the test set via spatial correlation, they are usually also not sufficient to train models that are able to generalize to other geographic areas or time points (e.g., different seasons). More modern datasets aim to increase diversity regarding scene content (e.g., more different object instances), environmental factors (e.g., seasons, light conditions, and so on), or aspects regarding image acquisition/processing (e.g., different look angles, resolutions, and so forth). This increase of diversity has thus far always connected to an increase in dataset size by including more images and/or larger scenes, or even other modalities. Although an increase in the provided image data is often easily possible, it is usually not feasible to have the same increase in the reference data. This leads to large-scale datasets where the reference data are much less carefully curated as for early datasets, often based on other existing sources (e.g., OpenStreetMap) and containing more label noise. Although many ML methods can handle a certain extent of label noise during training, its effects on the evaluation are barely understood and often ignored.

THE GEE DATA CATALOG CONTAINS MORE THAN 40 YEARS OF HISTORICAL IMAGERY AND SCIENTIFIC DATASETS FOR EARTH SCIENCE, WHICH ARE UPDATED AND EXPANDED DAILY.

CURRENT TRENDS: SPECIFICITY AND GENERALITY

In this context, two main characteristics of a dataset for a given task will be the focus of the discussion: specificity and generality.

EO applications present numerous and diverse scenarios due to varying object types, semantic classes, sensor types and modalities, spatial, spectral, and temporal resolutions, and coverage (global, regional, or local).

High specificity refers to datasets that are strongly tailored to a specific sensor-platform-task combination, maybe even being limited to certain acquisition modes, geographic regions, or meteorological seasons. These can hardly be used for anything beyond their original purpose. Although different application domains such as agriculture, urban mapping, military target detection, and bio-/geophysical parameter extraction do require different types of data, i.e., images, point measurements, tables, and metadata, the proliferation of different datasets specialized for every small task variation reduces reusability of the datasets in a broader context and affects scientific transparency and reproducibility.

High specificity also contributes to cluttered nomenclature, causing different datasets to appear different while

actually sharing very similar content. For example, a high level of detail in class semantics and terminology makes it difficult to compare the reference data of different datasets. A typical example is land cover classification, where similar classes may be aggregated into different subcategories depending on the application. As a result, models trained on different application-specific datasets may actually approximate very similar functional relationships between image data and target variables.

Virtually all of the less recent and still most of the modern datasets aim for specificity. However, several of the more recent benchmarks follow another direction: generality, i.e., providing more sensor modalities than actually required plus large-scale, often noisy reference data for multiple tasks instead of small scale and carefully curated annotations that only address a single task.

The contribution of such general datasets is manifold: first and foremost, the required number of (annotated) training samples for fully supervised ML simply does not scale very well given the effort of data annotation and curation in remote sensing. Thus, such general, large-scale datasets introduce new factors that increase the relation to realistic application scenarios such as robustness to label noise (e.g., by leveraging existing semantic maps as reference data, which are often outdated, misaligned, or of coarse resolution) and weakly supervised learning (where the reference data have a lower level of detail than the actual target variable, e.g., training semantic segmentation networks with labels on image level). Large-scale datasets are the only option to realistically test the generalization capabilities of learning-based models, e.g., over different geographic regions, seasons, or other domain

gaps. Furthermore, although multimodal datasets enable data-fusion and cross-modal approaches that leverage the different input sources, multitask datasets allow exploiting the mutual overlap of related tasks regarding feature extraction and representation learning. Finally, the idea of loosening the previously tight relationship between input data and the target variable in datasets (up to the point where a dataset might not offer reference data for any target variable) is to provide data that can be leveraged to learn powerful representations that are useful for a large variety of downstream tasks (as in pretraining or self-supervised learning).

However, there is not yet a single “go-to” dataset that can be used for pretraining most of the newly developed models or for benchmarking specific tasks against state-of-the-art approaches. Collecting such a high-quality benchmark dataset that enables pretraining of models for as many downstream tasks as possible is of significant value for further pushing performance boundaries.

Figure 28 presents a schematic diagram of the properties of an ideal solution for a go-to EO benchmark dataset, covering diverse geolocations, multiple modalities, different acquisition scenarios, and various applications. It is ideally acquired by different types of sensors and platforms with different viewing geometries to cover objects from different look angles. The images are obtained from different electromagnetic spectrum bands, i.e., visible, infrared, thermal, and microwave, resulting in multi-/hyperspectral, SAR, lidar, optical, thermal, and passive microwave measurements. The reference information or annotations are provided on a level that allows defining various tasks based on a single annotation. For example, an image with dense semantic annotations allows users to generate their desired object instance annotation files. Extending the dataset to multiple images of a scene with corresponding semantic labels enables not only semantic segmentation but also semantic CD tasks.

In summary, we foresee a certain duality in the future development of EO-related datasets: on the one hand, following the paradigm of data-centric ML [44], i.e., moving against the current trend of creating performance gains merely by leveraging more training data but instead focusing on datasets tailored toward specific problems with well-curated, high-quality reference data (e.g., manually annotated or based on official sources). On the other hand, general datasets that cover as many input and output modalities as possible to allow learning generic representations that are of value for a large number of possible downstream tasks.

FINDABILITY, ACCESSIBILITY, INTEROPERABILITY, REUSE AND ARD

In addition to the content, scope, and purpose of datasets, their organization will gain importance. With only a dozen public datasets available prior to 2015, it was feasible that each is provided with its own data format and meta-information, hosted on individual web pages, and

The Ideal Pretraining Dataset

A dataset ideally suited for pretraining and/or self-supervised learning should adhere to as many of the following characteristics as possible:

- ▶ multiple platforms (vehicle, drone, airplane, and satellite)
- ▶ multiple sensors (Planet, SPOT, WorldView, *Landsat*, *Sentinel* 1/2, and so forth)
- ▶ several acquisition modalities (SAR, RGB, hyperspectral, multispectral, thermal, lidar, passive microwave, and so on)
- ▶ diverse acquisition geometries (viewing angles, e.g., off-nadir conditions and spatial and temporal baselines in multiview data, e.g., interferometric SAR)
- ▶ realistic distortion factors (cloud cover, dust, smog, fog, atmospheric influence, spatial misalignments and temporal changes in multiview data, and so forth)
- ▶ well distributed geographical locations (spatial distribution within the dataset, climate zones, socioeconomic and cultural factors, and different topographies)
- ▶ diverse land cover/use (urban, rural, forest, agricultural, water, and so on)
- ▶ varying spatial resolution (0.1–1 m, 3–10 m, 10–30 m, 100–500 m, and scale distribution)
- ▶ temporally well distributed (seasonality, lighting condition, sun angle, and nighttime imagery)
- ▶ a diverse set of reference data that are well aligned with the EO measurements (semantic information, change, geo/biophysical parameters, and so forth).

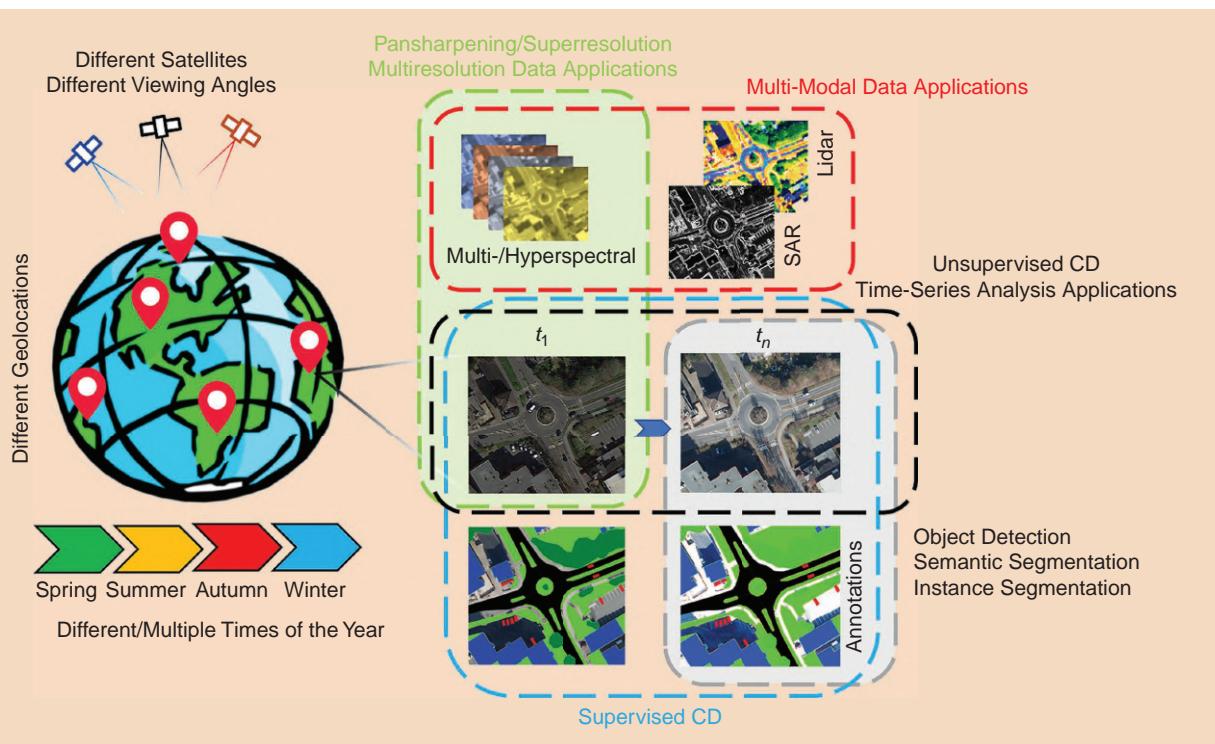


FIGURE 28. An illustration that shows the authors' view of the paramount properties that an ideal benchmark dataset needs to satisfy, including the type of tasks, sensors, temporal constraints, and geolocalization.

downloaded by everyone who wanted to work with them. With the hundreds of datasets available today and many more published every year, this cannot be maintained. Concepts such as findability, accessibility, interoperability, and reuse (FAIR) (see, for example, [45]) were proposed years ago and are still of high relevance. Data catalogs such as EOD (see the “Working With Remote Sensing Datasets” section) are a first step toward structuring datasets that are scattered among different data hosts. ARD (see, e.g., [46]), for example, in the form of data cubes [47], and efforts to homogenize meta-information, e.g., in the form of datasheets [48], will continue to evolve into new standardized data formats. The trend of datasets growing in terms of size and volume (see the “Evolution of EO-Oriented ML Datasets” section) as well as the need for global data products will soon put a stop to the current practice of downloading datasets and processing them locally. Working with data on distributed cloud services will create new requirements regarding data formats but also lead to new standards and best practices for processing. Finally, the goal of any ML-centered dataset is to train an ML model. These models should be treated similar to the data they originated from, i.e., they should follow standardized data formats and FAIR principles.

SUMMARY AND CONCLUSION

This article discussed the relevance of ML-oriented datasets in the technical disciplines of EO and remote sensing. An analysis of historical developments shows that the DL

boom has not only led to a rise in dataset numbers but also a large increase in size (as in spatial coverage and resolution but also with respect to multimodal and multitemporal imagery) and a diversity of application tasks. Furthermore, this development has led to the implementation of dedicated software packages and meta databases that help interested users develop solutions for their applications. Eventually, we drew the conclusion that one of the critical challenges in dataset design for EO tasks is the strong heterogeneity of possible sensors, data, and applications, which has led to a jungle of narrow-focused datasets. Although one of the trends in DL is certainly the consideration of smaller, well-curated, and task-specific datasets, another direction is the creation of a generic, nearly sensor- and task-agnostic database similar to the well-known ImageNet dataset used in Computer Vision. Such a generic dataset will be especially valuable in the pretraining of large high-capacity models with worldwide applicability.

AUTHOR INFORMATION

MICHAEL SCHMITT (michael.schmitt@unibw.de) received his Dipl.-Ing. degree in geodesy and geoinformation, his Dr.-Ing. Degree in remote sensing, and his habilitation in data fusion from the Technical University of Munich (TUM), Germany, in 2009, 2014, and 2018, respectively. Since 2021, he has been a full professor for Earth observation at the Department of Aerospace Engineering of the University of the Bundeswehr Munich, 85577 Neubiberg, Germany. Before that, he was a professor of applied geodesy

and remote sensing at the Department of Geoinformatics, Munich University of Applied Sciences. From 2015 to 2020, he was a senior researcher and deputy head at the Professorship for Signal Processing in Earth Observation at TUM; in 2019, he was additionally appointed adjunct teaching professor at the Department of Aerospace and Geodesy of TUM. In 2016, he was a guest scientist at the University of Massachusetts, Amherst. He is a cochair of the Active Microwave Remote Sensing Working Group of the International Society for Photogrammetry and Remote Sensing, and also of the Benchmarking Working Group of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee. He frequently serves as a reviewer for a number of renowned international journals and conferences and has received several Best Reviewer Awards. His research focuses on technical aspects of Earth observation, in particular image analysis and machine learning applied to the extraction of information from multisensor remote sensing observations. Among his core interests is remote sensing data fusion with a focus on synthetic aperture radar and optical data. He is a Senior Member of IEEE.

SEYED ALI AHMADI (cpt.ahmadisnpiol@yahoo.com) received his B.Sc. degree in surveying engineering and his M.Sc. degree in remote sensing from the Faculty of Geodesy and Geomatics, K.N. Toosi University of Technology, Tehran 19697, Iran, in 2015 and 2017, respectively, where he is currently pursuing his Ph.D. thesis on building damage assessment. He worked on image classification and segmentation techniques, machine learning algorithms, and lidar data processing. His thesis was focused on classifying hyperspectral and lidar datasets by combining spectral and spatial features to increase classification accuracy. He is a cochair of the Benchmarking Working Group of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee, frequently serves as a reviewer for a number of international journals, and received a Best Reviewer Award in 2018. His research interests include machine learning, deep learning, geospatial data analysis, image processing, and computer vision techniques for remote sensing and Earth observation applications.

YONGHAO XU (yonghaoxu@ieee.org) received his B.S. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, China, in 2016 and 2021, respectively. He is currently a postdoctoral researcher at the Institute of Advanced Research in Artificial Intelligence, 1030 Vienna, Austria. His research interests include remote sensing, computer vision, and machine learning. He is a Member of IEEE.

GÜLŞEN TAŞKIN (gulsen.taskin@itu.edu.tr) received her B.S. degree in geomatics engineering and her M.S. and Ph.D. degrees in computational science and engineering from Istanbul Technical University, Turkey, in 2001, 2003, and 2011, respectively. She is currently an associate professor at the Institute of Disaster Management

at Istanbul Technical University Istanbul, 34469 Turkey. She was a visiting scholar at the School of Electrical and Computer Engineering and the School of Civil Engineering at Purdue University from 2008 to 2009 and 2016 to 2017. She is a reviewer for *Photogrammetric Engineering and Remote Sensing*, *IEEE Transactions on Geoscience and Remote Sensing*, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *IEEE Geoscience and Remote Sensing Letters*, and *IEEE Transactions on Image Processing*. Her current research interests include machine learning approaches in hyperspectral image analysis, dimensionality reduction, explainable artificial intelligence, and sensitivity analysis.

UJJWAL VERMA (ujjwal.verma@manipal.edu) received his Ph.D. degree from Télécom ParisTech, University of Paris-Saclay, Paris, France, in image analysis and his M.S. degree (Research) from IMT Atlantique (France) in signal and image processing. He is currently an associate professor and head of the Department of Electronics and Communication Engineering at Manipal Institute of Technology, Bengaluru 560064, India. He is a recipient of the ISCA Young Scientist Award 2017–2018 by the Indian Science Congress Association, a professional body under the Department of Science and Technology, Government of India. He is also a recipient of the Young Professional Volunteer Award 2020 by the IEEE Mangalore Subsection in recognition of his outstanding contribution to IEEE activities. He is a co-lead of the Working Group on Machine/Deep Learning for Image Analysis of the Image Analysis and the IEEE Geoscience and Remote Sensing Society Data Fusion Technical Committee. He is a guest editor for Special Stream in *IEEE Geoscience and Remote Sensing Letters* and a reviewer for several journals including *IEEE Transactions on Image Processing*, *IEEE Transactions on Geoscience and Remote Sensing*, *IEEE Geoscience and Remote Sensing Letters*. He is also a sectional recorder for the Information and Communication Technology Section of the Indian Science Congress Association for 2020–2024. His research interests include computer vision and machine learning, focusing on variational methods in image segmentation, deep learning methods for scene understanding, and semantic segmentation of aerial images. He is a Senior Member of IEEE.

FRANCESCO PAOLO SICA (francescopaolo.sica@unibw.de) received his laurea (M.S.) degree (summa cum laude) in telecommunications engineering and his Dr. Ing. (Ph.D.) degree in information engineering from the University of Naples Federico II, Italy, in 2012 and 2016, respectively. Since 2022, he has been deputy head of the Earth Observation Laboratory at the Department of Aerospace Engineering of the University of the Bundeswehr Munich, 85577 Neubiberg, Germany. Between 2016 and 2022, he was a researcher at the German Aerospace Center. He received a Living Planet Post-Doctoral Fellowship from the European Space Agency for the High-Resolution Forest Coverage with InSAR & Deforestation Surveillance project. He is a cochair of the Benchmarking Working

Group of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee and a regular reviewer for international journals and conferences. His research interests cover a wide range of activities related to synthetic aperture radar (SAR) technology, from mission design to SAR signal and image processing, to end-user applications. He is a Member of IEEE.

RONNY HÄNSCH (ronny.haensch@dlr.de) received his diploma in computer science and his Ph.D. degree from TU Berlin, Germany, in 2007 and 2014, respectively. He is a scientist at the Microwave and Radar Institute of the German Aerospace Center, 82234 Weßling, Germany, where he leads the Machine Learning Team in the Signal Processing Group of the SAR Technology Department. He continues to lecture at TU Berlin in the Computer Vision and Remote Sensing Group. He serves as chair of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee, cochair of the ISPRS Working Group on Image Orientation and Fusion, IEEE Geoscience and Remote Sensing Society (GRSS) membership chair, organizer of the IEEE Conference on Computer Vision and Pattern Recognition Workshops with EarthVision (2017–2023), Photogrammetric Computer Vision (2019 and 2023), the Machine Learning for Remote Sensing Data Analysis Workshop at the International Conference on Learning Representations (2023), and the IEEE International Geoscience and Remote Sensing Symposium Tutorial on Machine Learning in Remote Sensing (2017–2023). He also serves as editor of *IEEE Geoscience and Remote Sensing Society eNewsletter* and associate editor of *Geoscience and Remote Sensing Letters* and *ISPRS Journal of Photogrammetry and Remote Sensing*. He has extensive experience in organizing remote sensing community competitions (e.g., the IEEE GRSS Data Fusion Contest 2018–2023), serves as the GRSS representative within SpaceNet, and was technical lead of the SpaceNet 8 Challenge. His research interest is computer vision and machine learning with a focus on remote sensing (in particular, synthetic aperture radar processing and analysis). He is a Senior Member of IEEE.

REFERENCES

- [1] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017, doi: 10.1109/JPROC.2017.2675998.
- [2] D. Hong, J. Hu, J. Yao, J. Chanussot, and X. X. Zhu, "Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 178, nos. 9–10, pp. 68–80, Aug. 2021, doi: 10.1016/j.isprsjprs.2021.05.011.
- [3] K. Li, G. Wan, G. Chen, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogrammetry Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020, doi: 10.1016/j.isprsjprs.2019.11.023.
- [4] Y. Long et al., "On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and million-AID," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4205–4230, Apr. 2021, doi: 10.1109/JSTARS.2021.3070368.
- [5] M. Schmitt, S. A. Ahmadi, and R. Hänsch, "There is no data like more data - Current status of machine learning datasets in remote sensing," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2021, pp. 1206–1209, doi: 10.1109/IGARSS47720.2021.9555129.
- [6] M. Schmitt, P. Ghamisi, N. Yokoya, and R. Hänsch, "EOD: The IEEE GRSS Earth observation database," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2022, pp. 5365–5368, doi: 10.1109/IGARSS46834.2022.9884725.
- [7] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279, doi: 10.1145/1869790.1869829.
- [8] F. Rottensteiner et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. I-3, no. 1, pp. 293–298, Sep. 2012, doi: 10.5194/isprsaannals-I-3-293-2012.
- [9] P. Ghamisi and N. Yokoya, "IMG2DSM: Height simulation from single imagery using conditional generative adversarial net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 794–798, May 2018, doi: 10.1109/LGRS.2018.2806945.
- [10] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, Jan. 2012, doi: 10.1109/TPAMI.2011.94.
- [11] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 4, no. 2, pp. 22–40, Jun. 2016, doi: 10.1109/MGRS.2016.2540798.
- [12] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag. (replaced Newsletter)*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.
- [13] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, Feb. 2020, Art. no. 111716, doi: 10.1016/j.rse.2020.111716.
- [14] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, "Functional map of the world," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6172–6180, doi: 10.1109/CVPR.2018.00646.
- [15] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: A meta-analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 309–322, Mar. 2021, doi: 10.1016/j.isprsjprs.2021.01.020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271621000265>
- [16] 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," *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 153–160, Sep. 2019, doi: 10.5194/isprsaannals-IV-2-W7-153-2019.
- [17] M. Schmitt and Y.-L. Wu, "Remote sensing image classification with the sen12ms dataset," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. V-2-2021, pp. 101–106, Apr. 2021, doi: 10.5194/isprsaannals-V-2-2021-101-2021.

- [18] F. Rottensteiner, G. Sohn, M. Gerke, and J. D. Wegner, "ISPRS test project on urban classification and 3D building reconstruction - 2D semantic labeling - Vaihingen data," Int. Soc. Photogrammetry Remote Sens., Hannover, Germany, 2013. Accessed: Oct. 14, 2022. [Online]. Available: <https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx>
- [19] M. Cramer, "The DGPF-Test on digital airborne camera evaluation - Overview and test design," *Photogrammetrie Fernerkundung Geoinf.*, vol. 2010, no. 2, pp. 73–82, May 2010, doi: 10.1127/1432-8364/2010/0041.
- [20] Y. Wang and M. Li, "Urban impervious surface detection from remote sensing images: A review of the methods and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 3, pp. 64–93, Sep. 2019, doi: 10.1109/MGRS.2019.2927260.
- [21] D. Wen et al., "Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 68–101, Dec. 2021, doi: 10.1109/MGRS.2021.3063465.
- [22] A. Karpatne, Z. Jiang, R. R. Vatsavai, S. Shekhar, and V. Kumar, "Monitoring land-cover changes: A machine-learning perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 8–21, Jun. 2016, doi: 10.1109/MGRS.2016.2528038.
- [23] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019, doi: 10.1109/JSTARS.2019.2918242.
- [24] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Introducing EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2018, pp. 204–207, doi: 10.1109/IGARSS.2018.8519248.
- [25] S. A. Yamashkin, A. A. Yamashkin, V. V. Zanozin, M. M. Radovanovic, and A. N. Barmin, "Improving the efficiency of deep learning methods in remote sensing data analysis: Geosystem approach," *IEEE Access*, vol. 8, pp. 179,516–179,529, Sep. 2020, doi: 10.1109/ACCESS.2020.3028030.
- [26] C. Broni-Bediako, Y. Murata, L. H. Mormille, and M. Atsumi, "Searching for CNN architectures for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2021.3097938.
- [27] G. Sumbul et al., "BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [Software and Data Sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 174–180, Sep. 2021, doi: 10.1109/MGRS.2021.3089174.
- [28] U. Chaudhuri, S. Dey, M. Datcu, B. Banerjee, and A. Bhattacharya, "Interband retrieval and classification using the multilabeled sentinel-2 BigEarthNet archive," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9884–9898, Sep. 2021, doi: 10.1109/JSTARS.2021.3112209.
- [29] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016, doi: 10.1016/j.isprsjprs.2016.03.014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0924271616300144>
- [30] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018.
- [31] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning ROI transformer for oriented object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2844–2853, doi: 10.1109/CVPR.2019.00296.
- [32] J. Ding et al., "Object detection in aerial images: A large-scale benchmark and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7778–7796, Nov. 2022, doi: 10.1109/TPAMI.2021.3117983.
- [33] H. Chen and Z. Shi, "A spatial-temporal attention-based method and a new dataset for remote sensing image change detection," *Remote Sens.*, vol. 12, no. 10, May 2020, Art. no. 1662, doi: 10.3390/rs12101662. [Online]. Available: <https://www.mdpi.com/2072-4292/12/10/1662>
- [34] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 2115–2118, doi: 10.1109/IGARSS.2018.8518015.
- [35] M. Märtens, D. Izzo, A. Krzic, and D. Cox, "Super-resolution of PROBA-v images using convolutional neural networks," 2019. [Online]. Available: <https://arxiv.org/abs/1907.01821>
- [36] G. Vivone, M. Dalla Mura, A. Garzelli, and F. Pacifici, "A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6102–6118, 2021, doi: 10.1109/JSTARS.2021.3086877.
- [37] J. Cornebise, I. Orsolic, and F. Kalaitzis, "Open high-resolution satellite imagery: The worldstrat dataset – With application to super-resolution," in *Proc. 36th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2022. [Online]. Available: <https://openreview.net/forum?id=DEigo9L8xZA>
- [38] Z. Xiong, F. Zhang, Y. Wang, Y. Shi, and X. X. Zhu, "Earth-Nets: Empowering AI in Earth observation," 2022, *arXiv:2210.04936*.
- [39] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Neural Inf. Process. Syst.*, 2019, vol. 32, pp. 8026–8037.
- [40] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX Symp. Oper. Syst. Des. Implementations*, 2016, pp. 265–283.
- [41] A. J. Stewart, C. Robinson, I. A. Corley, A. Ortiz, J. M. L. Ferres, and A. Banerjee, "TorchGeo: Deep learning with geospatial data," 2021, *arXiv:2111.08872*.
- [42] H. Tamiminia, B. Salehi, M. Mahdianpari, L. Quackenbush, S. Adeli, and B. Brisco, "Google earth engine for geo-big data applications: A meta-analysis and systematic review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 164, pp. 152–170, Jun. 2020, doi: 10.1016/j.isprsjprs.2020.04.001.
- [43] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "SpaceNet: A remote sensing dataset and challenge series," 2018, *arXiv:1807.01232*.
- [44] E. Strickland, "Andrew Ng, AI minimalist: The machine-learning pioneer says small is the new big," *IEEE Spectr.*, vol. 59, no. 4, pp. 22–50, Apr. 2022, doi: 10.1109/MSPEC.2022.9754503.

- [45] M. D. Wilkinson et al., "The fair guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, Mar. 2016, Art. no. 160018, doi: 10.1038/sdata.2016.18.
- [46] J. L. Dwyer, D. P. Roy, B. Sauer, C. B. Jenkerson, H. K. Zhang, and L. Lymburner, "Analysis ready data: Enabling analysis of the Landsat archive," *Remote Sens.*, vol. 10, no. 9, Aug. 2018, Art. no. 1363, doi: 10.3390/rs10091363. [Online]. Available: <https://www.mdpi.com/2072-4292/10/9/1363>
- [47] G. Giuliani et al., "Building an earth observations data cube: Lessons learned from the Swiss data cube (SDC) on generating analysis ready data (ARD)," *Big Earth Data*, vol. 1, nos. 1–2, pp. 100–117, Sep. 2017, doi: 10.1080/20964471.2017.1398903.
- [48] T. Gebru et al., "Datasheets for datasets," 2018. [Online]. Available: <https://arxiv.org/abs/1803.09010>
- [49] "SEN12MS toolbox." GitHub. Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/schmitt-muc/SEN12MS>
- [50] "EuroSAT: Land use and land cover classification with sentinel-2." GitHub Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/phelber/EuroSAT>
- [51] "Accurate and scalable processing of big data in earth observation." BigEarth. Accessed: Jul. 23, 2023. [Online]. Available: <https://bigearth.eu/>
- [52] "Functional map of the world (fMoW) dataset." GitHub. Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/fMoW/dataset>
- [53] "xView3: Dark vessels." Accessed: Jul. 23, 2023. [Online]. Available: <https://iuv.xview.us/>
- [54] "A large-scale benchmark and challenges for object detection in aerial images." DOTA. Accessed: Jul. 23, 2023. [Online]. Available: <https://captain-whu.github.io/DOTA/>
- [55] "LEVIR-CD." Accessed: Jul. 23, 2023. [Online]. Available: <https://justchenhao.github.io/LEVIR/>
- [56] "Onera satellite change detection dataset." GitHub. Accessed: Jul. 23, 2023. [Online]. Available: <https://rcdaudt.github.io/oscd/>
- [57] "Data," European Space Agency, Paris, France, Jun. 2019. [Online]. Available: <https://kelvins.esa.int/proba-v-super-resolution/data/>
- [58] "PAirMax-Airbus." Accessed: Jul. 23, 2023. [Online]. Available: <https://perscido.univ-grenoble-alpes.fr/datasets/DS353>
- [59] "The WorldStrat dataset." WorldStrat. Accessed: Jul. 23, 2023. [Online]. Available: <https://worldstrat.github.io/>
- [60] "Copernicus open access hub." Copernicus. Accessed: Jul. 23, 2023. [Online]. Available: <https://scihub.copernicus.eu/>
- [61] "Open data program." Maxar. Accessed: Jul. 23, 2023. [Online]. Available: <https://www.maxar.com/open-data>
- [62] "Capella space synthetic aperture radar (SAR) open dataset." AWS. Accessed: Jul. 23, 2023. [Online]. Available: https://registry.opendata.aws/capella_opendata/
- [63] "TerraSAR-X science service system." TerraSAR. Accessed: Jul. 23, 2023. [Online]. Available: <https://sss.terrasar-x.dlr.de/>
- [64] IEEE DataPort. Accessed: Jul. 23, 2023. [Online]. Available: <https://ieee-dataport.org/datasets>
- [65] "Open library for earth observations machine learning," Radiant Earth Foundation, Washington, DC, USA. Accessed: Jul. 23, 2023. [Online]. Available: <https://mlhub.earth>
- [66] "The earth in a cube." Euro Data Cube. Accessed: Jul. 23, 2023. [Online]. Available: <https://eurodatacube.com>
- [67] "Your gateway to NASA earth observation data," Earth Data, Nat. Aeronaut. Space Admin., Washington, DC, USA. Accessed: Jul. 23, 2023. [Online]. Available: <https://www.earthdata.nasa.gov>
- [68] "OpenAerialMap." Accessed: Jul. 23, 2023. [Online]. Available: <https://openaerialmap.org>
- [69] "OpenStreetMap." Accessed: Jul. 23, 2023. [Online]. Available: <https://openstreetmap.org>
- [70] "EarthNets for earth observation." EarthNets. Accessed: Jul. 23, 2023. [Online]. Available: <https://earthnets.nicepage.io>
- [71] "Earth observation database." Accessed: Jul. 23, 2023. [Online]. Available: <https://eod-grss-ieee.com>
- [72] "Torchgeo." GitHub. Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/microsoft/torchgeo>
- [73] "Azavea/raster-vision." GitHub. Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/azavea/raster-vision>
- [74] "Raster vision." Accessed: Jul. 23, 2023. [Online]. Available: <https://docs.rastervision.io/>
- [75] "Keras spatial." Github. Accessed: Jul. 23, 2023. [Online]. Available: <https://github.com/IllinoisStateGeologicalSurvey/keras-spatial>
- [76] "A planetary-scale platform for Earth science data & analysis." Google Earth Engine. Accessed: Jul. 23, 2023. [Online]. Available: <https://earthengine.google.com>
- [77] "NAIP: National agriculture imagery program." Earth Engine Data Catalog. Accessed: Jul. 23, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/USDA_NAIP_DOQQ
- [78] "Start building on AWS today." AWS. Accessed: Jul. 23, 2023. [Online]. Available: <https://aws.amazon.com>
- [79] "Welcome to Digital Earth Africa (DE Africa)." Digital Earth Africa. Accessed: Jul. 23, 2023. [Online]. Available: <https://www.digitalearthafrica.org>
- [80] "NOAA's emergency response imagery," National Ocean Service, Silver Spring, MD, USA, 2023. Accessed: Jul. 23, 2023. [Online]. Available: <https://oceanservice.noaa.gov/hazards/emergency-response-imagery.html>
- [81] "A planetary computer for a sustainable future." Planetary Computer. [Online]. Available: <https://planetarycomputer.microsoft.com>
- [82] "High resolution electricity access." Planetary Computer. Accessed: Jul. 23, 2023. [Online]. Available: <https://planetarycomputer.microsoft.com/dataset/hrea>
- [83] "USBuildingFootprints." GitHub. [Online]. Available: <https://github.com/Microsoft/USBuildingFootprints>
- [84] "Welcome to Colab!" Colab. Accessed: Jul. 23, 2023. [Online]. Available: <https://colab.research.google.com>
- [85] Accessed: Jul. 23, 2023. [Online]. Available: <https://www.grss-ieee.org/wp-content/uploads/2023/05/EODatasets.pdf>
- [86] "Start with more than a blinking cursor." Kaggle. [Online]. Available: <https://www.kaggle.com>
- [87] "EDC browser." Accessed: Jul. 23, 2023. [Online]. Available: <https://browser.eurodatacube.com/>