Deep Learning and Earth Observation to Support the Sustainable Development Goals

Current approaches, open challenges, and future opportunities

CLAUDIO PERSELLO, JAN DIRK WEGNER, RONNY HÄNSCH, DEVIS TUIA, PEDRAM GHAMISI, MILA KOEVA, AND GUSTAU CAMPS-VALLS

xxxxx

The synergistic combination of deep learning (DL) models and Earth observation (EO) promises significant advances to support the Sustainable Development Goals (SDGs). New developments and a plethora of applications are already changing the way humanity will face the challenges of our planet. This article reviews current DL approaches for EO data, along with their applications toward monitoring and achieving the SDGs most impacted by the rapid development of DL in EO. We systematically review case studies to achieve zero hunger, create sustainable cities, deliver tenure security, mitigate and adapt to climate change, and preserve biodiversity. Important societal, economic, and environmental implications are covered. Exciting times are coming when algorithms and Earth data can help in our endeavor to address the climate crisis and support more sustainable development.

THE EXPANDING ROLE OF DEEP LEARNING AND EARTH OBSERVATION FOR THE SUSTAINABLE DEVELOPMENT GOALS

Machine learning has played a fundamental role in the analysis of EO data for more than three decades, and its importance has been continuously growing. From the early investigations in artificial neural networks and statistical techniques [1], [2], the EO community has been striving for effective algorithms to automate the extraction of information from various sources of remotely sensed images, in situ data, and models. The developments in sensor technologies and the increasing availability of voluminous data go hand in hand with the demand for more accurate and scalable information extraction methods and tools. This demand is spurred by many geospatial applications and the growing awareness of the necessity to monitor system Earth for multiple threats to our natural environment, our climate, and the sustainable development of human societies.

After a long period when neural networks fell out of fashion, the DL revolution started about a decade ago and brought back attention to these powerful learning algorithms [3], [4]. Thanks to the development of specialized hardware, i.e., GPUs, and availability of large benchmark data sets, DL networks became more popular. They revealed extremely versatile learning machines able to learn virtually any task in data and image analysis [5]. Neural networks can be seen as trainable data-processing graphs, where the input data are gradually transformed through a sequence of layers that extract intermediate features and are finally used to predict the target output.

In a supervised setting, the network is trained with a set of input–output instances, exemplifying the functional relationship between the explanatory covariates and the target variable to predict. Although simplistic, this view depicts the flexibility of neural networks for data analysis purposes. A variety of architectures have been developed so far, the most popular being multilayer perceptrons [6], convolutional neural networks (CNNs) [7], recurrent neural networks (RNNs) [8], autoencoders [9], and generative adversarial networks [10]. Moreover, as the number of consecutive layers increases, i.e., the network becomes deeper, the algorithm tends to improve its ability to learn informative features, capturing intricate structures within input variables and their relations with the target output.

In the context of EO applications, deep networks can address a large variety of analysis tasks, from image classification and segmentation to data fusion, change detection, object detection, and delineation. Deep networks can be designed according to the characteristics of the remotely sensed data and, possibly, fuse different sensor data types and information layers. One of the main advantages of DL is the ability to learn abstract hierarchical representations of the data, allowing networks to bring spatial, spectral, and temporal patterns hidden in the data to the surface. This results in state-of-the-art performance and enables researchers and engineers to streamline the information extraction processing chain, potentially integrating multimodal data fusion, feature extraction, and inference tasks into one holistic, end-to-end learning framework. The combination of DL with powerful computing infrastructure and massive EO data sets opens up tremendous opportunities for geospatial applications. State-of-the-art DL methods are closing the gap between the performance of automated workflows and the need for accurate and reliable information imposed by real applications.

> EOs can generate data for monitoring a number of SDG targets and indicators. DL contributes to extracting meaningful and consistent information.

Moving beyond research laboratories, today, EO and DL have the opportunity to contribute to some of the most pressing global societal challenges, such as those identified by the 2030 Agenda for Sustainable Development [11]. The United Nations (UN) has defined a set of 17 SDGs as a plan of action to reach peace and prosperity for all people on our planet by 2030. The goals are related to social, economic, and environmental challenges, and they provide a blueprint for shared action. It is recognized that eradicating poverty in all its forms and dimensions is the greatest global challenge and an indispensable requirement for sustainable development. Each of the 17 goals has a set of targets and indicators to measure, monitor, and report the progress of each country.

The global framework established by the UN is designed around 169 targets and 232 indicators, representing the first truly data-driven framework in which countries can engage with evidence-based decision making and policy development [12]. The 2030 agenda recognizes that if you can't measure it, you can't manage it, thus emphasizing the importance of objective, accurate, and trustworthy information for decision making. This approach requires using multiple types of data, such as traditional national accounts, household surveys, and routine administrative data as well as new sources, such as EO data for the extraction of updated geospatial information.

The role of EO in support of the SDGs has been recognized and facilitated by international organizations, such as the Group on Earth Observation, Committee on Earth Observation Satellites, and European Space Agency (ESA) [12]–[14]. EO can provide continuous temporal information over the globe, capturing the sustainability of the developments underpinning the SDG framework. Satellite, airborne, and unmanned aerial vehicle (UAV) acquisitions provide data at multiple scales for monitoring the state of natural ecosystems, natural resources, oceans, coasts, land, and built infrastructure as well as their changes over time.

EO data are spatially and temporally consistent, allowing for effective comparisons of the results among different countries and in various years. EO data are also complementary with traditional statistical methods, offering a source of information to cross-check the validity of in situ data measurements (such as survey and inventory data) that are commonly collected by national statistical offices. Moreover, EO can significantly reduce the cost of monitoring SDG targets and indicators with respect to traditional data collection methods. According to the "Compendium of EO Contributions to the SDG Targets and Indicators" recently released by the ESA [14], 34 SDG indicators can be either directly (17 indicators) or indirectly (17 indicators) informed with space-based EO data across 29 targets and 11 goals.

Table 1 summarizes where EO data can contribute to SDG targets and indicators, providing examples of EO applications in support of monitoring the progress and achieving the goals. The ESA analysis [14] also recognizes the role of the technical infrastructure for storing and processing big EO data and, in particular, the relevance of cloud computing, parallel processing systems, and data cubes. However, the contribution of machine learning and DL toward the SDGs is not equally emphasized for their ability to extract meaningful and consistent information from EO data. This article aims to analyze the role of and opportunities for DL in EO to support the 2030 agenda for sustainable development (Figure 1).

DEEP LEARNING FOR EARTH OBSERVATION DATA

Advancements in DL, often based on computer vision research, had a large influence in EO image analysis, resulting in the adoption of DL for a variety of data types and geospatial applications [15]–[18].

TABLE 1. THE SDG TARGETS AND INDICATORS THAT CAN BE SUPPORTED BY EO-DERIVED PRODUCTS.

SDG	TARGETS	INDICATORS	EO APPLICATION IN SUPPORT OF SDG TARGETS AND INDICATORS
1 NO POVERTY	1.4	1.4.1, 1.4.2	Extraction of visible cadastral boundaries and information in support of fit-for-purpose land administration systems
	1.5		Risk assessment of natural and climate-induced disasters, early warning, and postevent damage assessment
2 ZERO HUNGER	2.3	2.3.1	Spatial distribution of cropland and smallholder farms as well as the estimation of agricultural productivity
	2.4	2.4.1	Assessment of the vulnerability to climate change, extreme weather, drought, flooding, and other disasters
3 GOOD HEALTH AND WELL-BEING	3.3		Early warning system for vector-borne disease
	3.6		Extraction of road maps and assessment of road conditions (paved/unpaved)
	3.9	3.9.1, 3.9.2	Mapping of hazardous chemicals and pollutants in the air, water, and soil
	3.d		Geospatial information in support of assessing health risks
5 GENDER EQUALITY	5.a	5.a.1	Extraction of visible cadastral boundaries and geospatial information in support of assessing ownership and securing rights over agricultural land
6 CLEAN WATER AND SANITATION	6.1, 6.3, 6.4	6.3.2	Mapping of water quality and pollutant concentrations
	6.5		Geospatial data for runoff modeling and global rainfall data
	6.6	6.6.1	Mapping of water-related ecosystems and changes in the extent of water-related ecosystems over time
7 AFFORDABLE AND CLEAN ENERGY	7.1	7.1.1	Mapping the human presence and availability of electricity (e.g., using nighttime images)
	7.2		Geospatial information in support of renewable energies
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE	9.1	9.1.1	Road and transportation network information in rural areas to support assessing accessibility to all-season roads

(Continued)

VERY-HIGH-RESOLUTION IMAGES

The analysis of very-high-resolution (VHR) images has been the first to benefit from DL. Given the large amount of spatial information and context contained in VHR images, the extraction of features has always been an active field of investigation [19]. With DL, it became possible to learn large dictionaries of convolutional filters directly from data. The appearance of publicly available large-scale data

TABLE 1. THE SDG TARGETS AND INDICATORS THAT CAN BE SUPPORTED BY EO-DERIVED PRODUCTS. (*Continued*)

11 SUSTAINABLE CITIES AND COMMUNITIES	11.1, 11.3	11.1.1, 11.3.1	Mapping of slum distributions and the extent, housing quality, density, and socioeconomic conditions of slum dwellers
	11.2	11.2.1	Road network information for assessing the accessibility to public transport
	11.4		Geospatial mapping and monitoring of cultural and natural heritage sites
	11.5		Risk assessment and early warning of vulnerable urban areas and disaster-induced damage assessment
	11.6	11.6.2	Air quality maps [particulate matter (PM)-2.5 and PM-10 concentration] and mapping waste sites
	11.7	11.7.1	Maps of urban green and public open spaces
	11.b, 11.c		Geospatial information for the development of resilient cities in developing countries
12 RESPONSIBLE CONSUMPTION AND PRODUCTION	12.2		Maps of natural resources
	12.4		Information about waste and pollutants released in the air, water, and soil
13 CLIMATE ACTION	13.1		Risks and damages associated with climate-related hazards and natural disasters
	13.2		Environmental variables for climate change models
14 LIFE BELOW WATER	14.1	14.1.1	Coastal eutrophication and floating plastic debris density
	14.2		Maps of marine and coastal ecosystems
	14.3	14.3.1	Marine acidity (pH)
	14.4	14.4.1	Geochemical (chlorophyll concentration) and geophysical analysis (sea surface temperature and ocean currents) and forecast for global and regional seas
15 LIFE ON LAND	15.1	15.1.1	Forest maps
	15.2	15.2.1	Forest inventories, deforestation/afforestation maps, and wildfire risk assessment
	15.3	15.3.1	Maps of deserts and degraded land as well as the prediction of drought and floods
	15.4	15.4.1;15.4.2	Mountain biodiversity maps
	15.5		Biodiversity maps
	15.7		Wildlife detection to support actions to end the poaching and trafficking of protected species

Our list differs slightly from those reported in [12]–[14]. We focus here on targets and indicators that can be more directly supported by EO applications and derived products and are not limited to satellite EO but also consider other EO platforms. (Source: https://www.un.org/sustainabledevelopment/) The content of this publication has not been approved by the United Nations and does not reflect the views of the United Nations or its officials or Member States.

sets issued from competitions enabled the appearance of deep networks specific for scene classification [20], [21] and the semantic segmentation [22]–[24] of VHR data. The authors in [20] introduced an explicit metric-learning regularization term in the loss function to learn more discriminative features. A large number of works appeared to process these data sets for classifying land cover at the single-pixel level: in [25], the authors proposed a hybrid system based on both CNN and traditional descriptor features and then used them in a random forest. In a subsequent article [26], the author trained two CNN models, one for color and the other for height data. In both cases, predictions were provided at the patch level (i.e., a single label was predicted for the whole patch), and a conditional random field was used to smooth the results.

After these first efforts, articles started to appear applying fully convolutional networks (FCNs) [27], providing predictions for each pixel of the patch in one go and greatly reducing the computational cost at the inference time [28]–[32]. Since these pioneering works, a large number of articles tackling semantic segmentation have been published in the field and pushed the boundaries of the performance on these data sets. Of notable interest are articles that tackled issues such as the integration of prior knowledge [33], edge information [34], and invariances [35] as well as the explicit inclusion of spatial reasoning [36].

Other works looked at methods to fuse the multiresolution bands acquired by most VHR satellite sensors, such as panchromatic (PAN) and multispectral (MS) images [37]–[39]. Bergado et al. [37], introduced a multiresolution
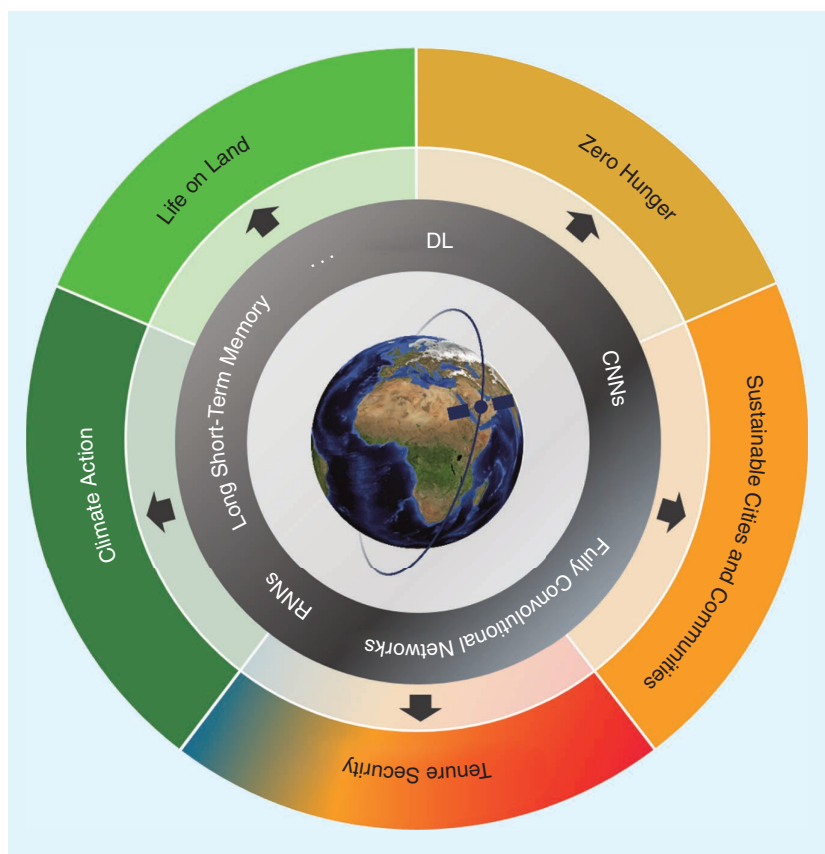
FIGURE 1. The framework of the study showing the considered application domains.
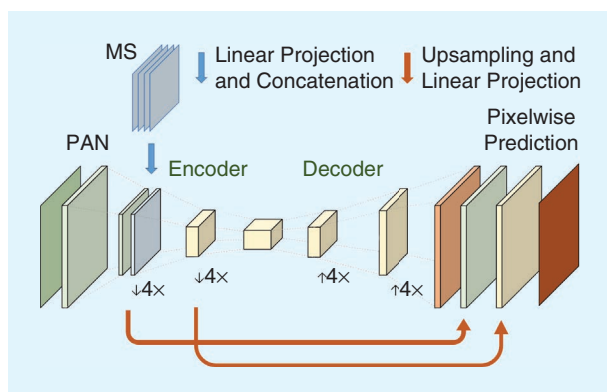

FIGURE 2. The architecture of the PAN–MS fusion network (FuseNet) for the pixelwise classification of VHR multiresolution images [37], [40].

FCN, called the *fusion network* (*FuseNet*), to perform an end-to-end image fusion and land cover classification (Figure 2). This architecture, tailored to VHR satellite data characteristics, resulted in higher performance than methods based on pan-sharpening. Contextual label information is included in ReuseNet [40], a fully convolutional recurrent network able to learn the contextual label-to-label dependencies that are commonly captured by techniques based on conditional random fields.

Going beyond the human design of the CNN architecture, Wang et al. [43] investigated a neural architecture search (NAS) approach [41], [42] to automatically design the CNN for the classification of VHR images. Unlike other NAS methods based on reinforcement learning or evolutionary algorithms over a discrete and nondifferentiable search space, their framework uses a gradient-based method to optimize both architecture and model parameters [44]. A switchable module allows for addressing both image classification and semantic segmentation.

Another promising research line investigates the extension of DL models to the direct prediction of regularized vector outcomes, i.e., outputs that can be immediately ingested in geographic information system (GIS) environments [45]–[48] (Figure 3). These developments are expected to have many practical uses for building outline delineation, road network extraction, and more, in general, for urban planning and monitoring applications in the context of SDG 11.

IMAGE TIME SERIES

RNNs are a powerful method for modeling sequential data, leading to much progress, especially in language processing [49]–[51]. Their capability of learning long-range patterns over time makes RNNs promising tools for a variety of tasks in remote sensing (RS), too. One important example in the context of SDGs is food security and the assessment of famine risk (SDG 2), which calls for large-scale mapping of agricultural activities. RNNs allow the learning of temporal patterns specific to different kinds of agricultural land use, as demonstrated in [52]–[55]. (See more details in the "Crop Type Mapping" section.)

A cell is the basic building block of an RNN. It combines the data of the current time step in the sequence and the unit's output from the previous time step as two inputs. While RNNs can, in principle, handle sequences of arbitrary and varying lengths, they are (in their basic form) challenged by long-term dependencies since learning those would require the propagation of gradients over many time steps. Gated architectures like long short-term memory (LSTM) cells [56] and gated recurrent units (GRUs) [57] aim at mitigating this problem. They use gating mechanisms to store and propagate information over longer time intervals to reduce the vanishing gradient problem. In general, abstract features are often represented better by deeper architectures [58]. In the same way that multiple hidden layers can be stacked in traditional feed-forward networks, multiple recurrent cells can also be stacked on top of each other; i.e., the output (or hidden state) of the lower cell is

connected to the input of the next-higher cell, allowing for different dynamics.

HYPERSPECTRAL IMAGES

Hyperspectral images (HSIs) have intensively contributed to SDGs, in particular, SDGs 2 [60], 6 [61], 14 [62], and 15 [63]. CNNs might be the most widely used deep architecture for feature extraction and classification due to the utilization of shared weights and local connections, which substantially decrease the number of trainable parameters in such networks compared to their fully connected alternatives. In the literature, 1D [64], 2D [65], and 3D CNNs [66] have been employed to extract spectral, spatial, and spectral–spatial information, respectively, from HSI images. The high dimensionality of HSIs, which leads to a greater number of trainable parameters compared to grayscale or MS images, along with the availability of only a limited number of training samples, make the training stage of such data extremely challenging. To address these issues, some basic strategies, such as dropout and weight decay, can be used. In addition, four sets of strategies have been investigated to properly train such high-dimensional data with only a limited number of training samples, such as dimensionality reduction [67], data augmentation [68], transfer learning [69], and semisupervised or even unsupervised learning [70].

RNNs have also been applied to HSI image analysis. By considering the spectral signature of each pixel vector as sequential data, an RNN can be applied to a single HSI for classification [71]. In this context, for each pixel vector, the spectral values are usually fed into the RNN from the first band to the last one (this can also be done in a bidirectional way [72]), and the output of the hidden layer at the last band is the extracted spectral feature. In real applications, the sequences' lengths can be very long (equal to the number of bands), which leads to training difficulties, such as gradient vanishing or explosion. To address this issue, a possible solution is to group the spectral bands into shorter sequences [73] or use LSTM [74] and GRU [75].

SYNTHETIC APERTURE RADAR

Synthetic aperture radar (SAR) emits coherent microwave pulses and records the amplitude and phase of their backscattered echo. As an active sensor, it is independent of daylight, and, due to the used frequencies, it can penetrate clouds; dust; and—to some degree—vegetation, soil, ice, and other materials. Applications range from estimating surface characteristics, such as roughness and moisture, to using polarimetric SAR for land cover/use classification, interferometric SAR (InSAR) for the generation of digital elevation models, and tomographic SAR to estimate height profiles over forests or urban areas.

However, SAR has not yet seen the same attention of DL as optical sensors. (For a detailed review of DL in SAR, we refer readers to [76].) The reasons for this are manifold. First, the imaging geometry differs greatly from optical cameras, causing effects unknown in optical imagery, such as layover or the displacement of moving objects. Furthermore, objects' appearances are strongly view dependent (e.g., certain types of backscatter happen only for certain geometric arrangements between the sensor and object). Second, SAR records the amplitude and phase of the received backscatter of a coherent pulse and is, therefore, complex valued.

While the absolute phase of a single-channel SAR image has no direct meaning, the relative phase between two polarimetric channels or two SAR acquisitions is highly important. Since most machine learning methods and frameworks are designed for real-valued data, early approaches to apply DL to SAR data relied on the extraction of real-valued (and mostly hand-crafted) features used as the input for the network (e.g., [77]). To address this issue, complex-valued CNNs and FCNs that directly work on the complex-valued data by using complex-valued convolutions and activation functions as well as loss functions are introduced in [78] and [79], respectively. An NAS approach is proposed in [80] to automate the CNN architecture design for SAR data and applied to a land cover/land use classification. CNNs for scene classification are studied in [81], and RNN architectures for object detection are investigated in [82].

The phase of multiple SAR images plays a particular role in InSAR as it relates to changes in height. It is used to generate digital elevation models as well as monitor earthquakes and volcanoes or general land subsidence. Corresponding networks need to be invariant to constant phase offsets and take the cyclic nature of the phase angle into account. CNNs have been used to enhance the quality of measured interferograms [83]; directly estimate the interferometric phase and coherence [84]; and perform phase unwrapping [85], i.e., the conversion of the cyclic phase into an absolute
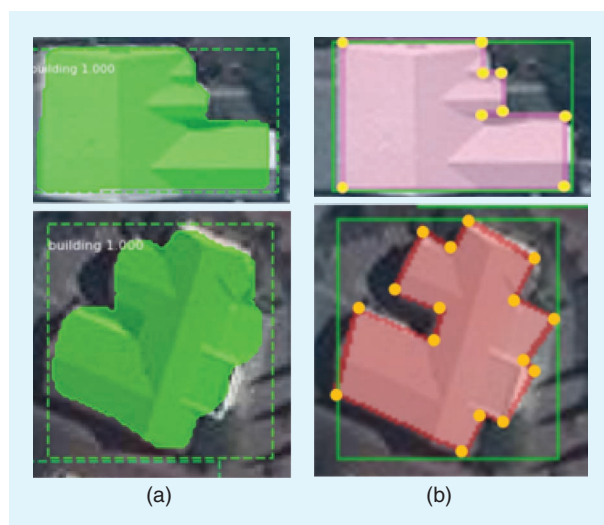

FIGURE 3. (a) The instance segmentation algorithm with raster output mask region-based CNN. (Source: [59]; used with permission.) (b) The regularized outline extraction with polygonal output. (Source: [48]; used with permission.)

phase field to estimate topographic heights or deformations. One particular problem in InSAR processing is decorrelation of the two SAR measurements, which can be due to several causes, including temporal changes and volumetric scattering. The latter often indicates vegetation and is used in [86] as input to a U-Net [87] to derive forests maps.

Another effect that has hindered the direct application of methods designed for optical images to SAR data is speckle: a chaotic fluctuation inherent to all measurements based on coherent waves caused by the interference of multiple backscattering in one resolution cell. Speckle reduction has greatly benefited from DL through supervised CNN-based denoising [88] [89] and a multistream complex-valued FCN [90] or by exploiting approaches, such as noise2noise (e.g., in [91] and [92]), that do not require clean data.

BIG GEODATA FUSION

A sharp increase in the amount of data captured by sensing devices has led to the big data deluge, creation of the new field of data science, and popularization of DL algorithms to deal with such data [93]. In a similar manner, the field of RS has been influenced by an ever-growing number of spaceborne, airborne, and proximate sensing devices, such as UAVs, to acquire multiscale data from a particular scene. The increase in the number, quality, and volume of passive sensing devices has been coupled with a growth in the number of alternative modes of measurement, such as airborne lidar, which generates point clouds representing elevation [94], and SAR sensors [95]. Furthermore, the new sources of ancillary data (e.g., data from crowdsourcing and social media [96]) have been used along with RS devices for a variety of applications in the context of smart cities and smart environments, hazard and disaster identification, and tracking.

Multisource data fusion aims to integrate the data of different types, distributions, and sources (they can be from a single sensor or different ones) by leveraging modality-specific information to improve the performance of the processing approaches compared to a single modality. CNN and its variants have significantly contributed to a wide range of multisource data fusion scenarios, such as the following:

◗ spatiospectral fusion to produce a fine-spectral-, fine-spatial-resolution image [97], [98]
◗ spatiotemporal fusion to create a fine-spatiotemporal-resolution image [99], [100]
◗ active (e.g., SAR and lidar) and passive (e.g., MS and hyperspectral) data fusion, mainly to improve classification performance or data matching [101], [102]
◗ RS and social media fusion [93], [103].

DEEP LEARNING APPLICATIONS CONTRIBUTING TO THE SUSTAINABLE DEVELOPMENT GOALS

This section provides an overview of DL and EO applications contributing to the monitoring and achievement of selected SDGs. We focus on applications in the context of zero hunger (SDG 2), sustainable cities and communities (SDG 11), tenure security (multiple SDGs), climate actions (SDG 13), and life on land (SDG 15).

ZERO HUNGER

Monitoring agricultural land use and production is essential to achieve zero hunger (SDG 2). It is of high importance for food production, biodiversity, and forestry [104]. An increasing world population, climate change, and changes in food consumption habits put yet-uncultivated areas under pressure while leading to intensification in existing agricultural areas [105]. Cropland expansion and the intensive use of agricultural areas are often connected with negative ecological impacts, like deforestation and biodiversity loss, but also the degradation of ecosystem services like ground and surface water quality [106], [107]. Therefore, the dense, accurate monitoring of agricultural lands plays an essential role for their optimal and sustainable management.

Some of the communities most vulnerable to hunger are smallholder farmers, who dominate agriculture in sub-Saharan Africa, with an estimated 51 million farms predominantly characterized by rain-fed production for household consumption [108]. African smallholder farmers often live in poverty in areas prone to natural hazards, where climate change is exacerbating the risks of hunger and breakdown of food systems. The large population growth in these areas urgently requires increased production, resilience to natural disasters, improvements in financial services, and the governance of food production systems. These improvements are fundamental for defeating hunger and malnutrition, realizing SDG 2 and, in particular, target 2.3, which aims to double the agricultural productivity and incomes of small-scale food producers by 2030. Sustainably increasing the productivity of these agricultural systems and, thus, improving food security and the livelihoods of smallholder families is a challenge, partly due to a lack of information about these systems.

The knowledge of crop areas and certain land uses is of importance for many political programs that aim to reduce and alleviate the environmental impacts of intensive agriculture, too [104]. Policy-driven incentives, for instance, encourage that a particular share of a farm's area remain extensively used grassland to promote biodiversity, or they give subsidies to promote a certain crop mix in the rotation [109]. Information collection is traditionally based on farmer self-reporting and spot-checking by authorities in the field, which is laborious, costly, and prone to errors.

Modern machine learning methods in combination with publicly available satellite imagery provide new possibilities for more accurate spatially dense monitoring of agricultural sites at high temporal resolution and low cost. One particularly promising recent sensor is *Sentinel-2*, due to its low ground sampling distance (10 m) at a revisit rate of three to five days. In general, the spectral signal of the

vegetation as captured by the satellite has specific characteristics as a function of the following:

◗ soil structure and composition (e.g., the soil brightness, water content, type, and so on)
◗ vegetation structure (e.g., the canopy cover, leaf area index (LAI), plant height, and leaf angle, among others)
◗ leaf biochemistry (e.g., chlorophyll, water and nitrogen content) [110].

Not only does each plant species have its own spectral signature, but spectral characteristics are also highly dependent on the phenological stage of the plant [111], [112]. Instead of merely analyzing images at a single point in time, time series (sequences) of satellite images provide significant additional evidence about crop species, and time-series analysis is a standard practice in agricultural RS.

CROP TYPE MAPPING

Crop classification from satellite data has been widely studied in RS. Traditional machine learning approaches with handcrafted features [113], [114] predominantly rely on vegetation indexes like the normalized difference vegetation index [115], [116]. Different strategies have been explored to include the temporal evolution as further evidence for classification, such as temporal windows [117], hidden Markov models and dynamic time warping [118], [119], and conditional random fields [120]. These traditional machine learning models have in common that they struggle to represent the complex spatiotemporal dynamics of spectral features.

> Machine learning in combination with open EO data provides new possibilities for monitoring agricultural sites at low cost.

Recent DL models no longer rely on hand-engineered features to encode spectral, spatial, and temporal patterns. They can learn very complex, highly nonlinear relationships if given sufficient labeled training data and computational resources. The authors in [122] propose the use of a CNN that combines cross-entropy and regression losses for simultaneously mapping and counting oil palm, coconut palm, and olive trees at the country scale. In [52], the authors use an RNN with LSTM to encode temporal dependencies in the data, while, in [123], the results on the same data set are improved by encoding both temporal and spatial dependencies via convolutional LSTM. In [124], satellite images are first processed individually with a CNN to obtain per-image features; then, temporal dependencies between these features are modeled with a separate RNN.

Further options are temporal CNNs that also combine features across time with convolutions [125] or models that use the attention principle [126] to aggregate information across time [53]. The work in [54] combines a pixel-set encoder and transformer [126] and shows an improved performance over RNN-based approaches. The authors in [55] build a deep RNN with a new cell structure termed *stackable recurrent (STAR)* that trains better than LSTM- and GRU-type models and is more parameter efficient. This makes it possible to train deeper models, which translates to improved performance across a range of sequence modeling tasks.

A recent alternative to RNN approaches for crop mapping involves neural ordinary differential equations, which can interpolate in the case of missing data [127]—due cloud coverage, for example. Finally, recent approaches have considered spatiotemporal bidirectional long short-term memory (bi-LSTM) architectures to fully exploit the information of long time series of high-resolution *Sentinel-2* data to classify different crop types (rice, fallow, barley, oat, wheat, sunflower, and triticale) [121] (see Figure 4).

DELINEATION OF FIELD BOUNDARIES

Field boundaries are essential for digital agricultural services enabling the estimation of cropland areas to aggregate and record specific information in a spatial database, such as the crop grown, soil type, yield, and application of pesticide and fertilizer. Moreover, they facilitate the extraction of land tenure boundaries for recording land rights in cadastral systems (see the "Deliver Tenure Security for All" section). Early research on field boundary delineation from EO data has focused on unsupervised techniques based on edge detection or segmentation [128]–[130]. These approaches are typically applied to areas characterized by intensive agriculture with large plots using medium-resolution images. However, small-area fields (<2 hectares) represent 40% of the fields worldwide and make up 70% of the cropland in Asia and Africa [131]. The delineation of such fields is extremely challenging since plots are small, irregularly shaped, and often with indistinct boundaries. In these circumstances, standard techniques fail in achieving the required accuracy.

To this end, DL-based strategies have resulted in significantly higher performance [132], [133]. An approach based on SegNet [134] and combinatorial grouping was proposed in [132] (Figure 5). The FCN is trained to detect field contours, discarding irrelevant edges. The detected sparse edges are then used as the input to the oriented watershed transform algorithm to extract a hierarchy of closed segments and iteratively merge adjacent regions based on the strength of their common boundary [135]. The final segmentation is obtained by applying the single-scale combinatorial grouping algorithm, which explores the segmentation hierarchy, to generate accurate field segments [136]. Promising results are obtained in two study areas in Nigeria and Mali. Marvaniya et al. [133] present a multistage approach that uses a combination of DL for edge detection and a sequence of postprocessing steps for improving the results.

Other recent DL-based solutions include [137]–[140]. A method based on U-Net and open data from the land parcel identification system of Spain was investigated in [137]. Waldner and Diakogiannis [138] adopted a multitask

approach to tackle the problem. They used ResUNet-a to identify the extent of fields, field boundaries, and distance to the closest boundary. Using a single monthly composite image from *Sentinel-2* as the input, their model could accurately map the field extent and boundaries. Other notable works have investigated a superresolution mapping approach [139] and the combination of neural networks with a graph-based growing contours method to extract agricultural field polygons [140].

SUSTAINABLE CITIES AND COMMUNITIES

Cities are the economic hubs of modern nations and home of an estimated 55.3% of the world's population. By 2030, urban areas are projected to house 60% of people globally, reaching 68% by 2050 [141]. While urbanization creates opportunities for economic developments, it also creates enormous social and environmental challenges. Some of the most pressing issues are the management of natural hazards, pollution, and the surge of socioeconomic inequalities resulting from excluding the poor from the social fabric.

> Approximately 1 billion people worldwide reside in informal settlements, living in deprived conditions and lacking access to essential services.

According to UN-Habitat, approximately 1 billion people worldwide reside in informal settlements, commonly called *slums*, living in deprived conditions and lacking access to essential services, such as safe water, acceptable sanitation, and durable housing [142]. The rapid urbanization processes in low- and middle-income countries contribute to the proliferation of deprived neighborhoods where dwellers live in crowded areas in unhealthy conditions and, often, without tenure security. In addition to that, these communities are also among the most vulnerable to the effects of climate change as well as the increasing frequency and intensity of natural disasters, such as floods, heatwaves, droughts, landslides, storms, wildfires, and cyclones [143].

MAPPING SLUMS AND URBAN POVERTY

The 2030 agenda pays particular attention to these global challenges with SDG 11, which aims at inclusive, safe, resilient, and sustainable cities and human settlements. The key indicator 11.1.1 requires monitoring "the proportion of urban population living in slums, informal settlements or inadequate housing." Current global statistics show a decline in the percentage of the urban population living in slums but an absolute increase of inhabitants living in such areas [144]. Nevertheless, official national statistics are often outdated, inconsistent, or simply inaccurate. Small slum pockets are generally neglected, and population counts based on census data are subject to large uncertainties, especially in large metropolitan areas [145]. More accurate and globally consistent methods to gather data on slum populations and their socioeconomic conditions are, therefore, needed.

Several studies show the ability of RS techniques to identify informal settlements, providing a relatively consistent mapping approach applicable over large areas and repeatable in time [146]. Maps derived from VHR satellite data can support SDG 11 and the monitoring indicator 11.1.1 in particular. Detailed 2D and 3D geospatial information extracted from UAV data can support the planning and monitoring of urban upgrading projects, thus contributing to targets 11.b and 11.c [147], [148] (Figure 6). The mapping of informal settlements can be performed on the basis of physical and morphological characteristics captured by VHR satellite images. Slums are commonly densely built-up areas characterized by small buildings arranged according to irregular layout patterns and a lack of green spaces.

Extracting these characteristics automatically from images is, however, a difficult task. The spectral information alone is insufficient to discriminate between different urban typologies (formal versus informal). It is necessary to extract contextual features capable of capturing long-range pixel dependency for distinguishing the different spatial patterns. Conventional machine learning approaches resort to the extraction of texture statistics, local binary patterns, oriented gradients, and segment-based features [149], [150]. However, these methods depend on several free parameters, which are difficult to optimize and, usually, set according to user experience.

The ability of CNNs to automatically learn high-level spatial features results in a streamlined workflow for slum mapping and higher classification accuracy. Mboga et al. [151] apply a CNN to detect informal settlements in Dar es Salaam, Tanzania, reporting an accuracy improvement over a support vector machine classifier trained with texture features and local binary patterns (Figure 7). The authors in [32] introduce FCNs for mapping informal settlements from VHR images. To this end, they adopt an FCN architecture with dilated convolutions [named *FCN-dilated kernel (DK)*], thereby capturing long-range pixel dependencies while keeping a limited number of network parameters (Figure 7). The best results are obtained by a network with six convolutional layers using increasing dilation factors. Moreover, they report a significant advantage in terms of the computational cost at testing time with respect to patch-based CNN.

Wurm et al. [152] investigate the transferability of an FCN model pretrained on VHR images to map slums in coarser-resolution *Sentinel-2* images and SAR data acquired by *TerraSAR-X*. They use an FCN-VGG19 architecture adapted from [27]. Their results show that transfer learning can significantly improve the results on *Sentinel-2* but not on *TerraSAR-X* data. Wang et al. [153] investigate a U-Net compound model, including dilated convolution operations, to map deprivation pockets in Bangalore, India. The authors in [154] use an FCN-based approach to study the temporal dynamics of slums, looking in particular, at temporary
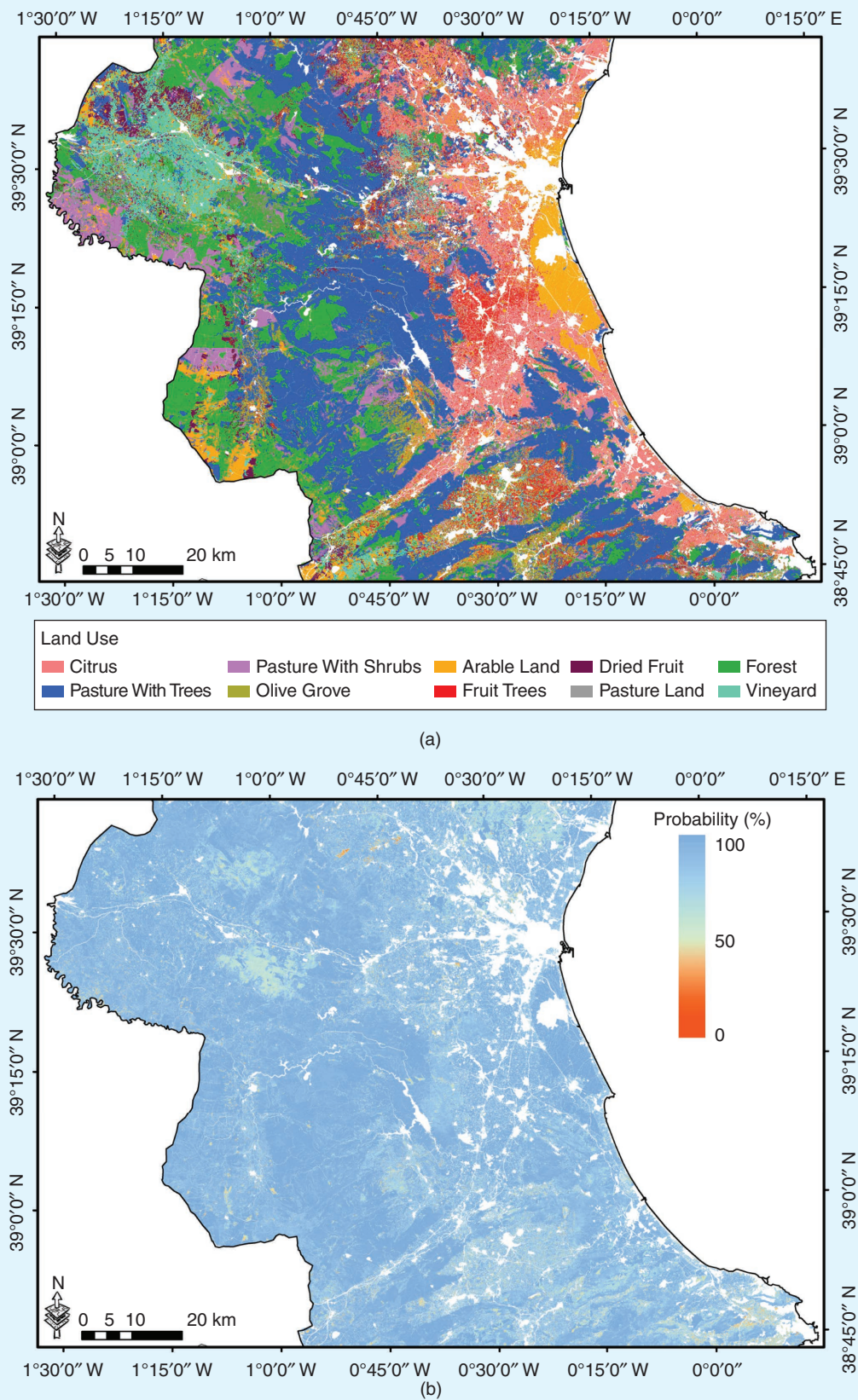
FIGURE 4. An example of spatiotemporal DL for crop monitoring from high-resolution images. A bi-LSTM deep network model was used to exploit the spatial regularities and temporal dimension of a sequence of *Sentinel-2* time series to predict the land use over (a) a number of classes of interest and (b) the pixel probability map. (Adapted from [121]; used with permission.)

slum pockets. The study investigates two change-detection approaches based on an FCN with dilated convolutions. The first approach uses a postclassification change detection, and the second trains FCNs to directly classify the transition in the land cover classes.

REVEALING SOCIOECONOMIC INEQUALITIES

The articles mentioned cast the slum mapping problem as a crisp classification, assuming that a boundary can be drawn to separate formal and deprived settlements. Departing from this dichotomy, Ajami et al. [155] adopt a framework conceptualizing the multidimensional nature of deprivation including not only the physical (e.g., poor house material) and financial levels (e.g., low-income residents), but also human, social, and contextual variables, such as accessibility to health care, education, and other services or social exclusion factors. The study introduces a data-driven approach to summarize multiple deprivation variables (both categorical and real valued) into a single

real-valued socioeconomic index, named the *data-driven index of multiple deprivations*. A CNN-based transfer-learning method predicts the socioeconomic index values based on VHR images and GIS features. The results show that an ensemble nonlinear regression model, combining the results of the CNN and models based on hand-crafted and GIS features, can explain 75% of the variation in the poverty index obtained from household data.

Other works have applied DL models to nighttime satellite images, street views, and aerial imagery to infer socioeconomic conditions. Jean et al. [156] use a CNN-based model to predict economic well-being across five African countries. The CNN model, pretrained on ImageNet, is fine-tuned to predict nighttime light intensities (used as a proxy for economic activities) corresponding to the input daytime satellite imagery. Finally, the CNN-extracted features, along with survey data, are used as the input to a ridge regression algorithm to infer the economic well-being. Social, environmental, and health conditions are extracted in [157] by
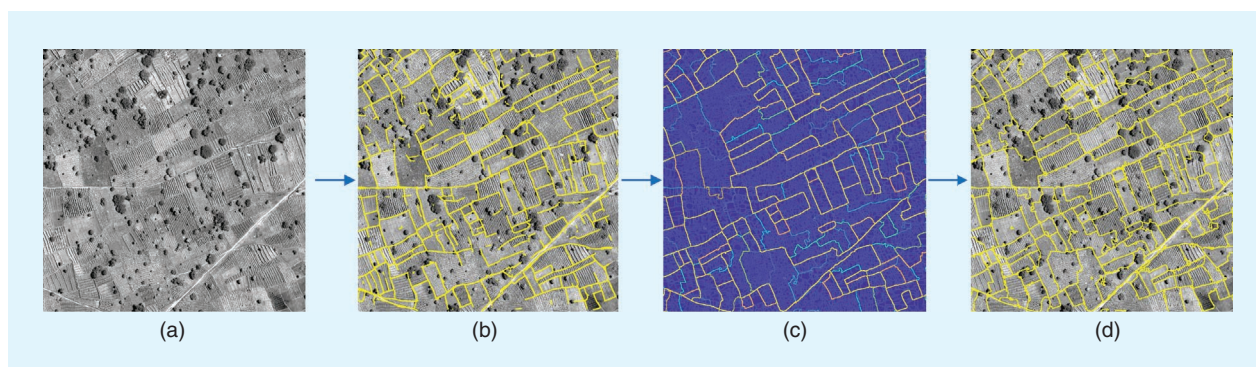

FIGURE 5. The DL workflow for the field boundary delineation in smallholder farms from (a) VHR imagery from *WorldView-3*. (Source: [132]; used with permission.) (b) SegNet is first applied to extract (fragmented) contours of the agricultural fields. (c) The oriented watershed transform is then utilized to extract a hierarchical segmentation. (d) Finally, the single-scale combinatorial grouping algorithm globalizes the local cues using an efficient implementation of normalized cuts and explores the combinatorial space of the segmentation hierarchy to generate regions that are likely to represent complete fields.

FIGURE 6. A UAV image acquired over an informal settlement in Kigali, Rwanda: the (a) 2D orthomosaic and (b) 3D photogrammetric point cloud. (Source: [147]; used with permission.)

a DL method applied to street-view images for major cities in the United Kingdom. Abitbol et al. [159] use a modified EfficientNetB0 CNN architecture [158] to predict socioeconomic status across France from aerial images and employ activation maps to interpret the urban topology.

DELIVER TENURE SECURITY FOR ALL

Secure property rights and efficient registration systems are essential for the modern economy. They give guaranty to individuals and businesses to invest in land, creating the conditions for improving the livelihoods and sustainable management of natural resources as well as enabling governments to collect property taxes, which are necessary to finance infrastructure and services for citizens. Unfortunately, a mere 30% of the global population has legally registered rights to their land and homes [160], [161], which means that more than 5 billion of the world's 7.8 billion people do not have documented land rights. Moreover, this percentage drops to 10% in African countries [162]. The insecurity of land tenure and property rights is often at the root of poverty and inequality [163], [164], leading to legal conflicts, unequal economic systems, and locks of assets, challenging effective and democratic governance principles.

The 2030 agenda recognizes the fundamental role of land rights security in several targets and indicators under SDGs 1, 2, 5, 11, 15, and 16. The correct registration of land tenure rights directly impacts food security, environmental sustainability, and the advance of women's empowerment worldwide. In many countries, the land is communally owned, but tenure insecurity is often the product of the government's inability to respond to the technical regularization needs [165].

> The insecurity of land tenure and property rights is often at the root of poverty and inequality.

Therefore, significant efforts are needed to formalize land ownership of the poor and vulnerable (target 1.4, indicator 1.4.2). Secure access to land is essential for small-scale agricultural producers to invest in their land and contribute to the market (targets 2.3 and 2.4). It is also fundamental for gender equality, ensuring women's rights to land tenure (SDG 5). The authors in [166] show that secure land tenure for women improves investments in agricultural developments and enhances the chance of women's involvement in family food and agricultural productivity. A lack of tenure security also impacts the development of sustainable cities (SDG 11), management of natural resources (SDG 15), and synergy between land administration agencies, courts, and legal support services (SDG 16).

Strategies to support these goals rely partly on the development of land administration systems (LASs) to formalize land rights and implement land-related policies [167]. There is a clear need for innovation for the fast, accurate, and cost-effective cadastral mapping needed for LASs [168]. The traditional surveying methods prove to be quite costly, slow, and labor intensive. In response,
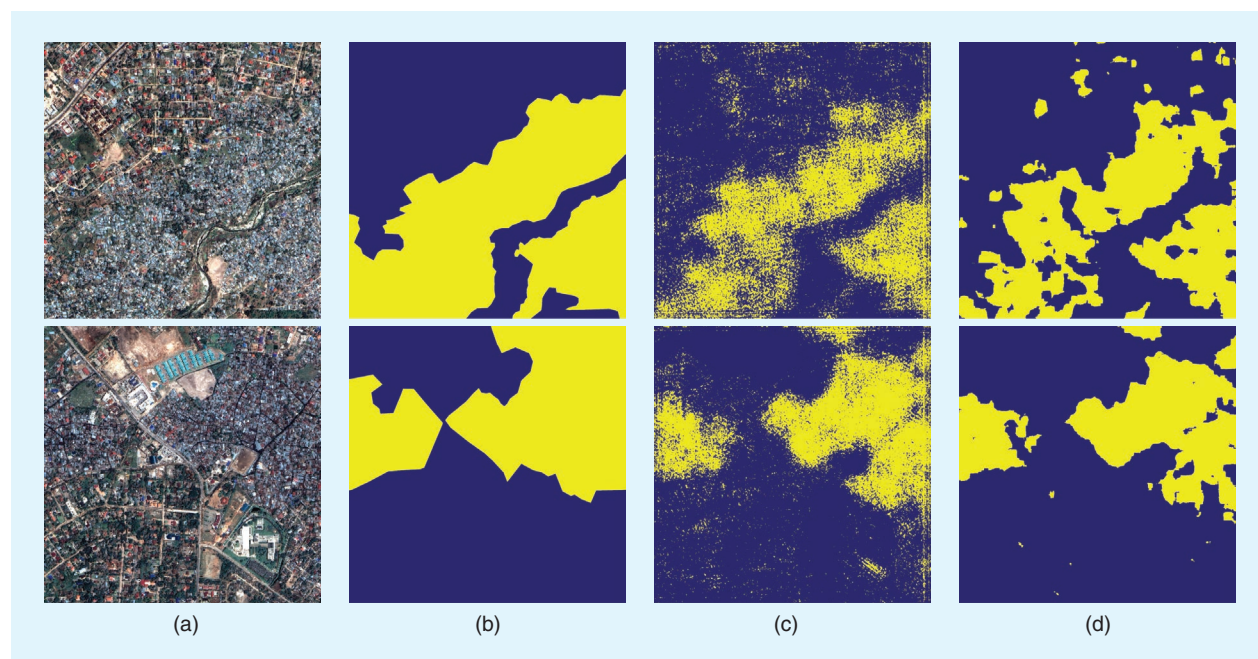

FIGURE 7. The informal settlement mapping over two test areas in Dar es Salaam, Tanzania, using a patch-based CNN and FCN-DK6: the (a) VHR image, (b) reference map, (c) patch-based CNN, and (d) FCN-DK6. "Informal settlements" are shown in yellow, and the "rest" are indicated by blue.

fit-for-purpose (FFP) land administration [169] advises and supports the development of new technologies using remotely sensed data and taking the country context into account. The FFP concept is also included in the recently developed framework for effective land administration developed by the UN Expert Group on Land Administration and Management, which is acting as a standard at the international level [170].

Spatial land rights recording, i.e., cadastral mapping, is the most expensive part of an LAS [171]. The automation or semiautomation of cadastral boundary delineation based on satellite or UAV images has been investigated since physical objects often coincide with visible cadastral boundaries and can be detected through image analysis [172]–[176]. The extracted boundaries should be combined with legal information, a procedure known as *adjudication*, and incorporate local knowledge from human operators.

The initially explored methods for the automated extraction of cadastral boundaries from EO images are based on image segmentation and edge detection [173]. However, the main disadvantages of these methods are sensitivity to intraparcel variability and the dependence of the selected parameters [177], [178]. Better results have been achieved using learning-based contour detectors, such as the globalized probability of boundary (gPb) [135], which combines brightness, color, and texture cues into a globalization framework using spectral clustering.

Recent studies explored DL methods [179], [180]. The consortium of the its4land project (https://its4land.com/) developed a suite of open source solutions for land tenure recording using EO data [168], [181]–[183]. It developed methods based on the gPb, simple linear iterative clustering superpixels, and a CNN to extract cadastral boundaries and a strategy to assign costs to each line incorporating local user knowledge. This work resulted in an open source plug-in for a Quantum Geographic Information System providing user-guided delineation functions calculating the least-cost paths along the extracted and weighted boundaries. Experiments were conducted using aerial images acquired in Ethiopia and UAV images from Rwanda and Kenya (Figure 8). Overall, the obtained results based on CNN-derived boundaries achieved a precision of 76%. The use of this semiautomated interactive method leads users to spend 38% less time and 80% fewer clicks compared to manual delineation [179], [184].

Following this research line, Xia et al. [180] explored the potential of FCNs to extract cadastral boundaries in urban and semiurban areas in Rwanda using UAV data. The authors adopted the FCN-DK architecture, which, compared to gPb and multiresolution segmentation, resulted in overall better performance. Nevertheless, the performance of any automated method depends on the presence of visible objects delimiting the boundary of the property (e.g., fences, pathways, walls, roads, or land cover transitions).

In support of the full land recording process, Chipofya et al. [185] developed an approach incorporating hand-drawn sketch maps with remotely sensed data. Their method converts the raster sketch map into a vector automatically, and the hand-drawn symbols are detected and recognized using a CNN. The system performs a stroke-based image segmentation wherein the boundaries of sketched objects are drawn and delineated. Finally, the concepts corresponding to the detected symbols are applied to the image segments based on the distance and a fixed set of rules specifying the spatial constraints on configurations of different types of features.

CLIMATE ACTION

The current scenario of climate change and projections from climate models call for definite and urgent action, as requested by SDG 13 [186], [187]. Extreme events are more severe, frequent, and unexpected in space and time [188]. The Earth system is actually changing globally but also on the local and regional scales, with huge implications for ecosystems, biodiversity, and agriculture, just to name a few. In this scenario, humanity faces the challenges of both the mitigation of and adaptation to climate change; that is, to try to reduce emissions as much as possible while preparing for unavoidable consequences that are no longer a future but a reality [189]. Machine learning and DL, in particular, can help in the myriad of aspects concerned in both issues. The 2030 agenda focuses mainly on adaptation aspects, with target 13.1 demanding the strengthening of resilience and the adaptive capacity to climate-related hazards and natural disasters in all countries.

The mitigation of greenhouse gas emissions requires important changes to electricity systems, transportation, buildings, industry, and land use. Adaptation requires planning for resilience and disaster management, given an understanding of climate and extreme events; see [189] for an organized collective effort to synthesize both the methods and challenges.

DEEP LEARNING FOR CLIMATE CHANGE MITIGATION

Reducing emissions can be achieved with machine learning and DL models. For instance, several DL models have been used to forecast electricity supply and demand, e.g., to create short- and medium-term forecasts of solar [190] and wind power [191], [192] or even use deep networks to produce demand forecasts that optimize for electricity scheduling costs rather than forecast accuracy [193]. DL in combination with RS satellite imagery has also been used to generate size and location data for rooftop solar panels [194], [195], and there are some deep networks that estimate the state of the system [196], [197]. As electricity gets transported from generators to consumers, some of it gets lost as resistive heat on electricity lines. Prior work has performed predictive maintenance using LSTMs [198] and neural network-plus-clustering techniques [199] on electric grid data.

Another important field of action is transportation. Decarbonizing transport is essential to a low-carbon society,

and there are numerous applications where machine learning can make an impact. For instance, vehicles can be detected in VHR images accurately [200]–[202], and image counts can serve to estimate the average vehicle traffic [203]. Neural networks have also been used for analyzing the preferences of customers traveling by high-speed trains.

Many critical systems inside buildings can be made radically more efficient. Deep autoencoders can be used to simplify information about machine operation so that deep neural networks can then more easily predict multiple kinds of faults [204]. Occupancy detection in buildings can help identify energy demands, a problem where deep neural networks have been also applied [205]. DL can also help to monitor and optimize the operations in smart buildings [206]. Machine learning may be able to assist with many aspects of CO2 sequestration. While still in its infancy, we have seen recent approaches on the use of convolutional image-to-image regression techniques for uncertainty quantification in a global carbon storage simulation study [207]. Such models can help in the development of novel strategies to monitor and develop underground carbon sequestration techniques.

DEEP LEARNING FOR CLIMATE CHANGE ADAPTATION

We use general circulation and Earth system models to anticipate climate scenarios on our planet as well as inform local and national governments for decision making. Models have become very precise in projecting scenarios, but they still disagree in some particular cases and are very computationally expensive to run. Machine learning in general and DL in particular can help to mitigate both aspects. The largest part of the uncertainty comes from the parameterization of clouds and aerosols in the models, which have clear implications, as bright clouds block sunlight and cool Earth. Deep neural networks have been used to emulate the behavior of high-resolution clouds, resolving simulations at a fraction of the computational cost [208]. Improvements are expected with the combination of DL and process understanding in a new form of hybrid modeling approaches that are data driven while respecting the fundamental laws of physics [4], [209]. Future improvements in climate modeling will necessarily have to account for the proper characterization and modeling of ice sheet dynamics and sea level rise, yet machine learning has not yet approached such problems systematically [209]–[211].

Weather models are optimized to track the rapid, chaotic changes of the atmosphere, and DL has recently impacted the associated problems. For instance, deep networks are now heavily used to make local forecasts from coarse 10–100-km climate or weather model predictions [212]; other researchers try to translate high-resolution climate forecasts into risk scenarios, e.g., of localized flooding patterns from past data [213], which have clear impacts on individuals. Accurately forecasting hazards and their impacts has societal, economical, and environmental implications. DL is now present in initiatives involving preserving ecosystems at risk [214]; monitoring the chances of food insecurity [215]; and deploying a swift, effective disaster response [216].

However, humans can also intervene in the system directly. This is the field of geoengineering. For example, neural network approaches could facilitate the fast release of aerosols in both space and time [217]. Modeling impacts is also of high relevance; the authors in [218] use deep neural networks to estimate the effects of aerosols on human health, while Crane-Droesch et al. [219] use them to estimate the effects of solar geoengineering on agriculture. Finally, we should note that geoengineering raises many ethical questions, where explainable, accountable artificial intelligence (AI) and fair learning should be part of the discussion.

LIFE ON LAND

SDG 15 aims to protect, restore, and promote the sustainable use of terrestrial ecosystems; sustainably manage forests; combat desertification; halt and reverse land degradation; and stop the loss of biodiversity. Achieving this goal has far-reaching consequences that are closely interlinked with many other SDGs. Here, we review several application domains where DL and EO play a central role.

SUSTAINABLE FOREST MANAGEMENT

The relevance of the sustainable management of forests is linked directly to SDG 15, and target 15.2 in particular, but it goes well beyond that. Indeed, the "The State of the World's Forests 2018" report of the Food and Agriculture Organization [220] identifies that forests and trees are relevant for 28 targets from 10 different SDGs. On the one hand, forests are a key variable to mitigate the effects of climate change; they protect the soil and water and contain more than 75% of the world's terrestrial biodiversity. On the other hand, forests provide products and services, such as food, medicine, and fuel, that are of high socioeconomic
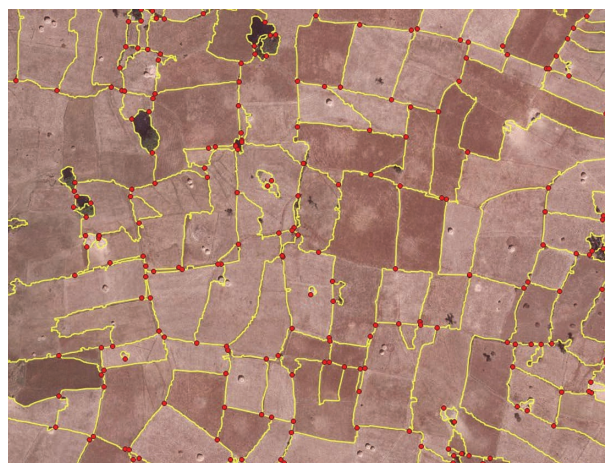

FIGURE 8. The cadastral mapping tool (its4land) applied to a rural area in Ethiopia.

importance in particular in rural areas. The combination of RS and DL has been extensively used to monitor forests, e.g., by producing global forest maps [86], delineating individual tree crowns in aerial imagery [221], [222], or performing damage assessment after storms [223].

Despite their relevance, the loss of the world's forests through deforestation and forest degradation is an increasing issue destroying natural habitats; limiting resources for the world's poorest; and, in the long term, worsening climate change by significantly contributing to CO_2 emissions. Reasons for deforestation include tree logging for materials, mining, and farming. Agriculture producing palm oil, beef, soy, pulp, and paper is responsible for nearly three quarters of the tropical deforestation [220].

> RS is the ideal tool for monitoring large forests that are difficult to access.

Deforestation is mostly happening in rural areas and often performed in secrecy, which requires monitoring large regions that are difficult to access. Thus, RS is the ideal tool for mapping and monitoring forests that inspired the publication of open data (e.g., in the context of the "Understanding the Amazon From Space" challenge organized by Planet [224]) as well as the usage of DL approaches. Several works focus on experimental comparisons between different network models and shallow learners [225], [226] or apply ensembles of different CNN architectures (e.g., [227]—a participant in the aforementioned challenge). In [228], deforestation mapping is modeled via spatiotemporal deep CNNs by taking several domain-specific components (e.g., the handling of clouds) into account. Modern approaches go beyond a mere mapping of forest areas or direct deforestation and, instead, aim for identifying possible reasons for forest loss. ForestNet [229] not only proposes a deep convolutional network to characterize the processes leading to deforestation but also provides a data set based on *Landsat 8* imagery of forest loss events annotated by expert interpreters.

WILDFIRE RISK

Wildfires, as one of the major factors contributing to deforestation, are becoming more frequent and destructive due to several reasons, including higher temperatures, increased droughts, fuel accumulation, and dead vegetation as well as increased population density in close proximity to forests and wildlands. Traditionally, wildfires are detected by human observers either by chance (and then reported to local emergency numbers) or from dedicated watchtowers. Current works aim to complement or even replace the latter by deploying ground-based camera networks (e.g., a High Performance Wireless Research and Education Network and Alert Wildfire in the state of California, United States, where, in 2018, more than 8,000 wildfires burned 800,000 hectares of land). The camera feed of these networks can then be automatically analyzed by DL approaches as, e.g.,

in [230], which uses an inception network to detect smoke. An alternative is the usage of UAVs as proposed in [231], which uses a saliency-based system to generate image region proposals that are then analyzed by a standard CNN for classification.

Satellite imagery can be used for wildfire detection as well but comes with its own challenges. Geosynchronous satellites, such as *GOES 16* or *GOES 17*, constantly observe large parts of a hemisphere but have a rather coarse resolution of several square kilometers, which makes the detection of wildfires in their early stages difficult. Nevertheless, their image data have been used in combination with DL for wildfire detection (e.g., in [232]). Orbiting satellites, such as *Moderate Resolution Imaging Spectroradiometer (MODIS)*, *Visible Infrared Imaging Radiometer Suite*, *Landsat*, and *Sentinel-1/-2 (VIIRS)*, on the other hand, have a much finer spatial resolution but revisit times of several hours to days. In particular, SAR sensors offer unique benefits, as they are able to penetrate clouds and smoke and are independent of daylight. *Sentinel-1* time-series data and DL have, for example, been used in [233] to provide a near-real-time progression monitoring of wildfires.

Beyond the detection and monitoring of wildfires, forecasting their future burns and spread is another important application area [234], [235]. Bergado et al. [235] use a big geodata set to predict wildfire burns. They design FCNs for predicting daily maps of the probability of a wildfire burn over the next seven days, utilizing an extensive set of wildfire-related input variables taken from various data sources. A total of 29 quantitative features are selected as the input to the models. These features encode factors associated with wildfire burn, such as topography (the elevation, slope, and aspect), weather (the temperature, humidity, solar radiation, rainfall, wind speed and direction, and lightning flash density), proximity to anthropogenic interfaces (the distance to and power lines), and fuel characteristics (the fuel type, fuel moisture, and emissivity). Historical wildfire burn records for Victoria, Australia, collected over the period of 2006–2017 are used for training and testing the DL models. DL and RS have also been used for postevent analysis, e.g., for damage assessment [236] or to analyze the impact of wildfires on tree species [237].

BIOPHYSICAL PARAMETER ESTIMATION

The problem of retrieving biogeophysical parameters spans a wide variety of applications and has a direct impact on achieving the SDGs. The related goals require creating spatially explicit and temporally resolved maps of quantities and essential climate variables to monitor vegetation status and health, agricultural, and forest production. Parameters should be estimated in a consistent and standardized manner to improve accountability.

The use and abuse of vegetation indexes as proxies for vegetation status and health has been challenged recently by machine learning approaches, from the nonlinear generalization of indexes [238] to more advanced, yet

supervised, machine learning models [209]. Many parameters are now estimated using machine learning; for example, surface temperature and moisture are key parameters for weather prediction, with great impacts on agriculture and the environment, in ecology, hydrology, meteorology, and biology, while the LAI and fractional vegetation cover help in assessing the vegetation cover and dynamics, with implications for crop production. Several seminal works relevant for DL parameter retrieval have been published with a focus on earth sciences [4], [209], environmental applications [16], and RS [15].

Land parameter retrieval often concerns biochemical parameters but can also include physical ones, such as the land surface temperature, which was retrieved from microwave radiometer data with DL in [241] and tested on reference data from both ground stations and other optical satellite data with good results. The LAI and leaf chlorophyll content have been retrieved with optical sensors and using neural networks, yet mostly using shallow architectures [242].

Retrieving parameters can often be hampered by the scarcity of measurements and observations to spatialize them with machine learning. This is the situation with the relevant parameters for monitoring the land and vegetation, such as the canopy water content. For such a case, one can resort to radiative transfer models to generate a lookup table of expressive simulations to learn from and upscale it in Google Earth Engine globally [243].

There are some other cases where samples are available in big databases but not sufficiently complete, with many missing attributes or uncertainty in the wild. This was the case for important leaf and plant traits, like phosphorus or nitrogen concentrations, that were not upscaled until the exploitation of the TRY database along with multisensor fusion and machine learning [244]. The upscaling of carbon, heat, and energy fluxes from eddy covariance data has been recently tackled with all kind of machine learning models and neural networks in particular. The key parameters for sensing the health and sensitivity of our warming planet are the gross primary production and net ecosystem exchange. Their estimation with neural networks and ensemble methods allow us to quantify global land–atmosphere interactions and benchmark land surface model simulations [245], [246].

Research in farming applications also relates to biological parameter retrieval. Often, though, the goal is not to use predictions as parameters in models but as proxies for the health condition of crops in so-called smart farming applications. By monitoring and optimizing these vegetation indexes, the goal is to increase the crop yield. The variables of interest, such as the crop type, crop yield, soil moisture, and weather variables, can also be used to model and understand the ecosystems that farming affects [247]. Most often, though, they are applied to data sets covering only smaller regions of agricultural areas. As opposed to biological parameter retrieval applications, DL is frequently used in farming applications. Some country-level work on agriculture has been done for, e.g., corn crop yield [248] and wheat [249], but little research exists on larger-scale studies where predictions could be used in models. The authors in [250] provide a comparison of several AI methods on a case study in the midwestern United States.

Forest cover, biomass, and vegetation height are other types of biological parameters that are of high importance to understand and monitor Earth, with obvious societal and economical implications. DL has also been applied to this problem, although mostly on the continent-level scale; e.g., in [251], researchers used the LSTM networks; the authors in [252] modeled forest dynamics over a 28-year period by stacking time series and formulating the task as a change-classification problem; and [253] predicted the above-ground forest biomass from lidar and *Landsat 8* data with stacked sparse autoencoders. The authors in [254] map the vegetation height densely at a 10-m resolution from stacks of *Sentinel-2* MS optical satellite imagery at the country scale using CNNs with a regression loss.

WILDLIFE CONSERVATION

A global loss of biodiversity is observed at all levels [255], and mammals are no exception, with one fifth of them at risk of extinction [256]. Conservation relies heavily on monitoring to estimate the biodiversity as well as resources to sustain life and risks related to human activities (hunting, poaching, expanding agriculture, and so on). Despite the urgency of protecting animal populations, the monitoring of them is more often done locally in reserves or by experts on foot and hardly meets the scaling and update requirements to monitor fauna effectively. Due to their larger field of view and the relatively high revisit time potential, satellites [257] and—more importantly—drones [258] are considered more for surveying by wildlife ecologists [259], [260]. Drones open perspectives for monitoring on demand, the safe detection of poachers, and an estimation of grazing potential. To process the sheer amount of data collected by drones, researchers are starting to resort to DL massively to identify animals in the wild with object detection pipelines [261]–[264].

These efforts go hand in hand with computer vision-based community efforts aiming at processing the wide archives of camera trap images, i.e., static cameras placed at strategic locations in reserves [265]. To support research in DL-based animal conservation, a number of software suites are being proposed, including Annotation Interface for Data-Driven Ecology (AIDE) [266], which allows ecologists to upload their camera traps or aerial survey and deploy pretrained (or their own) models in the cloud on Microsoft

RS and DL can accelerate conservation efforts and play a central role in the battle against poaching.

Azure. Examples of deployments of AIDE in camera trap image classification and single-animal detection are reported in Figure 9. The questions of accuracy with respect to animal size and image resolution or when related to the ratio of the background versus animal occupancy in the images (the animals only occupy a fraction of the data collected) are central in these articles. The detection and tracking of poachers is also on the rise, with approaches using thermal images at night [267] or based on deep reinforcement learning [268].

In the context of animal censuses, a question of wide interest is the time efficiency versus number of animals these algorithms miss. Precise counts are of prime importance, and a low recall would force rangers to go through the entire image collection for verification, which would negate the benefits of the DL detection pipeline. Recent research compared DL and citizen science counting methods [269] and concluded that both led to similar accuracy, with a significant speedup achieved when using DL. Also, an active research field is the joining of these two worlds via active learning [270] algorithms: by allowing an interactive back and forth between the annotators and DL models, significant speedups as well as increased generalization to new campaigns and transfer to new reserves have been achieved [271]. Finally, these interactive pipelines are currently made accessible to the large ecological community, for instance, via web-based platforms enabling interactive annotation guided by DL models classifying or detecting in the background [266].

CHALLENGES AND FUTURE OPPORTUNITIES
The previous section shows several examples of geospatial applications where DL and EO allow a systematic investigation of global phenomena, providing continuous and spatially consistent information supporting evidence-based decision making and local interventions. We expect that the coming decade will see a surge of research in this direction, with innovative methodological developments and an increase in the number and scope of applications in support of the SDGs.

However, several questions remain to be addressed: some are purely scientific, and others are at the interface among scientific communities, stakeholders, and decision makers. Will the EO scientific community succeed in producing accurate, reliable, consistent, and up-to-date geospatial information? Moreover, are these results trustworthy for governmental authorities, stakeholders, and local communities? In other words, are DL models trusted by nonexperts, who are in charge of decision making and policy development? The success of evidence-based decision making largely depends on the trust that people have in the data. Transparent data analysis methods and clear communication are fundamental to set proper expectations and build trust between data providers and decision makers.

OPEN CHALLENGES

UNCERTAINTY QUANTIFICATION
To be of true value and gain trust by people, DL models need to provide an indication of the reliability of model predictions. Assigning well-calibrated uncertainties to model outputs plays a critical role in many real-world applications. A significant additional benefit of the uncertainty estimates assigned to each data point of the model output is that this creates a practical interface to more traditional postprocessing steps using Bayesian models at their core.

We can define the *uncertainty* within DL in a twofold way: that inherent to all models is epistemic uncertainty, and that inherent to all data is aleatoric uncertainty. The former captures the dissimilarity of unseen data compared to what our model has been trained on, i.e., samples that lie within the training distribution have a low epistemic uncertainty, and those that are out of distribution have a high epistemic uncertainty. Aleatoric uncertainty results from the noise inherent in the observations, such as sensor noise [272]. In addition, there are uncertainties in the spatial domain that arise, for instance, when variables are aggregated over spatial units (e.g., districts or administrative units), resulting in the so-called modifiable areal unit problem. It is, thus, important to consider all sources of uncertainties and their propagation through the whole processing pipeline that affect the quality of the final product.

DATA QUALITY QUANTIFICATION FOR DECISION MAKING
It is essential to realize that, to support evidence-based policy making and promote data-driven planning and decision making, the quality of data products must be carefully assessed and carefully communicated so that nonexperts can understand. On the one hand, we encourage the scientific community to pay more attention to how data quality is assessed and communicated. On the other hand, we recommend that policy developers incorporate data uncertainties into the decision-making process explicitly. In this respect, we recognize the importance of defining standard data quality measures. Thus, we promote a tighter collaboration between the scientific community and policy makers to define standards on quality measures to quantify the SDG indicators.

MODEL EXPLAINABILITY
In addition to the uncertainty quantification, there is a growing interest in making machine learning and DL models more interpretable and understandable, aiming at neural networks that provide human-understandable justifications for their output, leading to insights about the inner workings [273], [274]. In EO, explainable AI is a relatively new field but quickly becoming important due to the implications that trustable black-box models can have on the usage of DL in societal applications.

In the context of agricultural EO, Campos-Taberner et al. [121] investigated how to deepen the understanding of an RNN for land use classification based on *Sentinel-2* time series. In [275], the authors studied how land use can be employed to explain the automatic prediction of the landscape's scenic value (a form of a cultural ecosystem service). To do so, they used semantic bottlenecks [276] as the intermediate layers of a regression network, predicting landscape beauty from *Sentinel-2* images. Forcing the network to choose among human-interpretable solutions, then recombined linearly, the model allows an understanding of why (in terms of land use) the model predicts a given score.

MODEL TRANSFERABILITY

One of the greatest challenges of DL in EO is the often-limited model transferability. For example, a slum-mapping DL model trained in Dar es Salaam is unlikely to produce accurate results in Bangalore or São Paulo. This happens not only because the RS images may be affected by different acquisition and radiometric conditions but also because cities in various parts of the world have different characteristics and definitions of what constitutes a slum. Despite several studies in domain adaptation and transfer learning [277], model transferability remains a challenge to ensure the spatial and national consistency of indicators derived from DL models. Moreover, nonexperts might be unaware of this problem. It is, therefore, essential that DL model developers provide clear guidance to users regarding the domain where the model is expected to produce valid results.

INTERDISCIPLINARY APPROACH

Addressing global societal problems requires a vast palette of expertise ranging from RS, DL algorithm development, and advanced computational skills as well as domain knowledge in fields such as agriculture, forestry, ecology, urban management and planning, social sciences, land administration, animal conservation, and so on. It requires researchers to collaborate and co-design solutions together with other scientists and engage with stakeholders, industrial partners, local communities, and governmental and nongovernmental organizations. The barriers between different scientific (and nonscientific) communities are often a challenge for an effective interdisciplinary approach.

FUTURE OPPORTUNITIES

In this article, we recognize the importance of the availability of data and computational facilities for the success of DL models. Developments in this direction are offering new opportunities to the EO community. For half a century, Earth has been under continuous observation by satellites to monitor and understand environmental processes. However, historically, RS data were foremost available to those governmental agencies, research institutes, and commercial companies that had direct access to the corresponding sensors. A mixture of different political, organizational,

and legal reasons made a free distribution of acquired data difficult to impossible, resulting in a limited number of mostly small data sets [278]. The traditional approach to developing and testing new methods on small and local data still prevails today. In particular, in the context of DL, this is problematic, as approaches are evaluated on data sets

Classification of Camera Traps Images
(a)

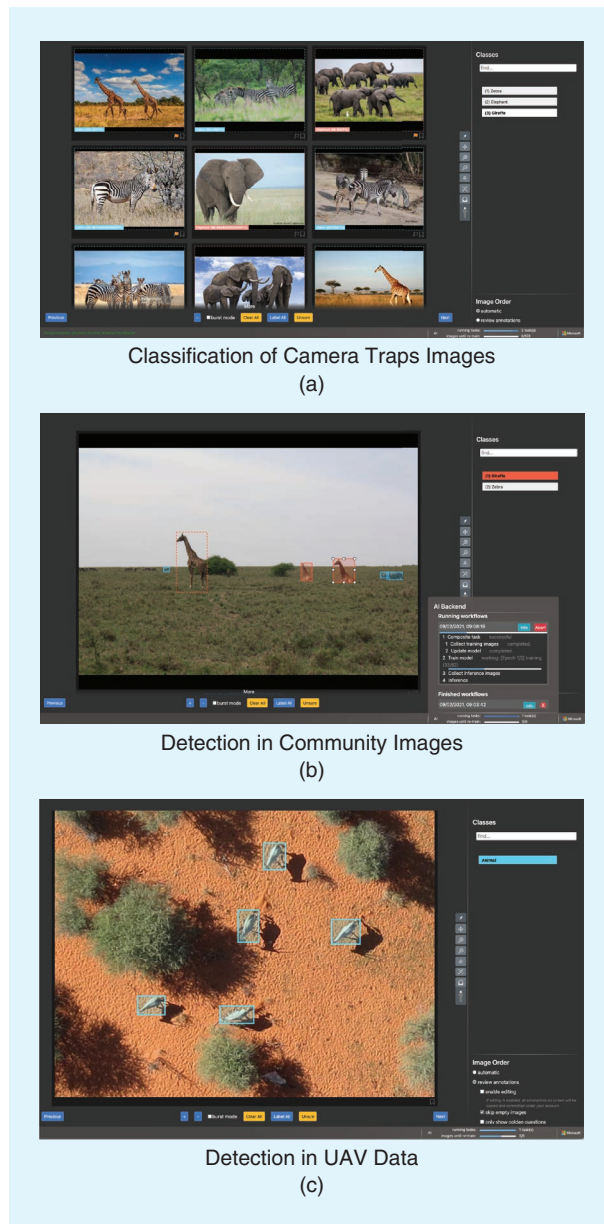Detection in Community Images
(b)

Detection in UAV Data
(c)

FIGURE 9. Three deployment cases of the AIDE platform [266] in camera trap images: for (a) camera trap image classification, (b) animal detection from tourists' and photographers' pictures acquired during a safari [using the "Great Zebra and Giraffe Count" (GZGC) campaign (http://lila.science/datasets/great-zebra-giraffe-id], and (c) detecting wildlife in UAV images in the Kuzikus reserve in Namibia. In the GZGC case, a typical deployment is shown: a user is editing the detections, while a model is training in the cloud (bottom right box); the predictions of the current model (dashed lines) are also used as guidance.

(often consisting of only a single small image as, e.g., the HSI data set Indian Pines) that do not provide a sufficient amount of independent test samples.

BIG AND OPEN GEODATA

Fortunately, during the last years, large parts of the community have been moving away from closed data and embraced open science principles such as findable, accessible, interoperable, and reusable [279] and free and open source software [280]. These developments enable transparent and reproducible scientific research, allow the distribution and reuse of data and methods, and lead to the more efficient creation of new data products. This led to open code libraries [such as Open RS (http://openremotesensing.net) and the IEEE RS Code Library (https://rscl-grss.org/), public evaluation servers [such as the IEEE Geoscience and Remote Sensing Society Data and Algorithm Standard Evaluation website (http://dase.grss-ieee.org)], and modern benchmark data sets for a multitude of combinations of EO sensors and tasks, which are rapidly replacing older small-scale data sets. These are more in line with the actual situation in RS, as current EO data have passed the petabyte scale and pose common big data challenges regarding volume (i.e., the amount of data), velocity (that is, the temporal pace at which new data are acquired), and variety (the heterogeneity regarding image acquisition, such as sensors types and modes as well as environmental factors) [281].

CLOUD COMPUTING INFRASTRUCTURES

New opportunities are also arising from the availability of cloud computing infrastructures that allow the visualization and analysis of large-scale data [e.g., Landsat [282] and Sentinel data (https://scihub.copernicus.eu/)] directly in the cloud without the need for local downloading, storing, and processing. Examples include Digital Earth Australia [283], Earth System Data Lab [284], Swiss Data Cube [285], Copernicus Data and Information Access Services, and Google Earth Engine [239]. The availability of large and open data sets in combination with powerful computing infrastructures sets the premise for researchers to work more cohesively on addressing the environmental and societal challenges of our time.

A GLOBAL PICTURE OF WORLDWIDE PHENOMENA

Finally, we want to remark that the combination of DL and EO offers the opportunity to obtain a truly global picture of environmental and societal phenomena that go beyond national boundaries as opposed to the data typically collected by national statistical agencies. As discussed in [240], the national-level reporting structure of the SDGs limits the ability to capture environmental phenomena that cross national borders. Moreover, differences in the data collection practices of the national offices often result in inconsistent data. EO can provide spatially and temporally consistent data, while DL offers the tools to extract semantic information in an objective and reproducible manner. We, thus, advocate for the use of DL and EO to monitor progress toward the SDGs and encourage the geoscience and RS communities to play active roles in discussions with stakeholders and policy makers.

CONCLUSION

We have reviewed the latest developments in the context of DL for EO and a large number of applications that contribute to the UN agenda for sustainable development. The combination of DL and EO appears to be a strategic asset that can play an essential role in addressing many of the challenges raised by the UN agenda and, beyond that, some of the most urgent demands of human societies. Understanding the role of DL in EO for extracting nationwide geospatial statistical data has far-reaching societal implications for policy development and decision making. Going beyond the SDG agenda, DL and EO can play significant roles in other international agendas, such as the New Urban Agenda (https://habitat3.org/the-new-urban-agenda) or Sendai framework for disaster risk reduction.

ACKNOWLEDGMENTS

AUTHOR INFORMATION

Claudio Persello (c.persello@utwente.nl) is an associate professor with the Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, 7500 AE, The Netherlands.

Jan Dirk Wegner (jandirk.wegner@uzh.ch) is with the Institute of Computational Science, University of Zurich, Zurich, 8057, Switzerland.

Ronny Hänsch (ronny.haensch@dlr.de) is with the SAR Technology Department, Microwaves and Radar Institute, German Aerospace Center, Wessling, 82234, Germany.

Devis Tuia (devis.tuia@epfl.ch) is with Ecole Polytechnique, Fédérale de Lausanne, Sion, CH-1951, Switzerland.

Pedram Ghamisi (p.ghamisi@gmail.com) is with Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Germany, and the Institute of Advanced Research in Artificial Intelligence, the AI4RS group, Vienna, 1030, Austria.

Mila Koeva (m.n.koeva@utwente.nl) is an associate professor with the Faculty of Geo-Information Science and

Earth Observation, University of Twente, Enschede, 7500 AE, The Netherlands.

Gustau Camps-Valls (gustau.camps@uv.es) is with the Universitat de València, Paterna, València, 46980, Spain.

REFERENCES

[1] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 540–552, 1990. doi: 10.1109/TGRS.1990.572944.

[2] J. Lee, R. C. Weger, S. K. Sengupta, and R. M. Welch, "A neural network approach to cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 5, pp. 846–855, 1990. doi: 10.1109/36.58972.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012. pp. 1–9.

[4] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and, Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[5] Y. LeCun, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[7] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 11 1997, doi: 10.1162/neco.1997.9.8.1735.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Sci. (New York, N.Y.)*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[10] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[11] "Transforming our world: the 2030 Agenda for Sustainable Development," United Nations General Assembly, 2015. [Online]. Available: https://sdgs.un.org/2030agenda

[12] European Space Agency (ESA). "Satellite earth observations in support of the sustainable development goals," in The CEOS Earth Observation Handbook, Special 2018 *ed.*, 2018, pp. 1–107.

[13] "Earth observations in support of the 2030 agenda for sustainable development," Group on Earth Observation (GEO), Tech. Rep., 2017. [Online]. Available: https://earthobservations.org/documents/publications/201703_geo_eo_for_2030_agenda.pdf

[14] "Earth observation for SDG: Compendium of earth observation contributions to the SDG targets and indicators," European Space Agency (ESA), 2020. https://eo4society.esa.int/wp-content/uploads/2021/01/EO_Compendium-for-SDGs.pdf

[15] X. X. Zhu et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[16] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, p. 111716, Jan. 2020, doi: 10.1016/j.rse.2020.111716.

[17] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, Jun. 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[18] T. Hoeser and C. Kuenzer, "Object detection and image segmentation with deep learning on earth observation data: A review-part I: Evolution and recent trends," *Remote Sens.*, vol. 12, no. 10, p. 1667, May 2020, doi: 10.3390/rs12101667.

[19] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, 2013, doi: 10.1109/JPROC.2012.2197589.

[20] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, 2018, doi: 10.1109/TGRS.2017.2783902.

[21] G. Cheng, X. Xie, J. Han, S. Member, and L. Guo, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020, doi: 10.1109/JSTARS.2020.3005403.

[22] F. Rottensteiner, G. Sohn, M. Gerke, J. D. Wegner, U. Breitkopf, and J. Jung, "Results of the ISPRS benchmark on urban object detection and 3d building reconstruction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 93, pp. 256–271, Jul. 2014, doi: 10.1016/j.isprsjprs.2013.10.004.

[23] 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: 2D contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5547–5559, 2016, doi: 10.1109/JSTARS.2016.2569162.

[24] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2017, pp. 3226–3229, doi: 10.1109/IGARSS.2017.8127684.

[25] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. Van Den Hengel, "Semantic labeling of aerial and satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2868–2881, 2016, doi: 10.1109/JSTARS.2016.2582921.

[26] J. Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," 2016, arXiv:1606.02585.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. Comput. Vision Pattern Recogn.*, Jun. 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[28] N. Audebert, B. L. Saux, and S. Lefèvre, "Semantic segmentation of earth observation data using multimodal and multi-scale

deep networks," in *Proc. Asian Conf. Comput. Vision (ACCV)*, 2016, pp. 180–196.

[29] M. Volpi and D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, 2017, doi: 10.1109/TGRS.2016.2616585.

[30] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, 2017, doi: 10.1109/TGRS.2016.2612821.

[31] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "High-resolution aerial image labeling with convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7092–7103, 2017, doi: 10.1109/TGRS.2017.2740362.

[32] C. Persello and A. Stein, "Deep fully convolutional networks for the detection of informal settlements in VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2325–2329, Dec. 2017, doi: 10.1109/LGRS.2017.2763738.

[33] D. Tuia, M. Volpi, and G. Moser, "Decision fusion with multiple spatial supports by conditional random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3277–3289, 2018, doi: 10.1109/TGRS.2018.2797316.

[34] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 158–172, Jan. 2018, doi: 10.1016/j.isprsjprs.2017.11.009.

[35] D. Marcos, M. Volpi, B. Kellenberger, and D. Tuia, "Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models," *ISPRS J. Int. Soc. Photo. Remote Sens.*, vol. 145, pp. 96–107, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.01.021.

[36] L. Mou, Y. Hua, and X. X. Zhu, "Relation matters: Relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 1–13, 2020, doi: 10.1109/TGRS.2020.2979552.

[37] J. R. Bergado, C. Persello, and A. Stein, "FuseNet: End-to-end multispectral VHR image fusion and classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2018, pp. 2091–2094, doi: 10.1109/IGARSS.2018.8519214.

[38] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, 2018, doi: 10.1109/JSTARS.2018.2805923.

[39] R. Gaetano, D. Ienco, K. Ose, and R. Cresson, "A two-branch CNN architecture for land cover classification of PAN and MS imagery," *Remote Sens.*, vol. 10, no. 11, p. 1746, Nov. 2018, doi: 10.3390/rs10111746.

[40] J. Bergado, C. Persello, and A. Stein, "Recurrent multiresolution convolutional networks for VHR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6361–6374, 2018, doi: 10.1109/TGRS.2018.2837357.

[41] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in *Proc. Int. Conf. Learning Representations*, 2017, pp. 1–16.

[42] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.

[43] J. Wang, Y. Zhong, Z. Zheng, A. Ma, and L. Zhang, "RSNet: The search for remote sensing deep neural networks in recognition tasks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2520–2534, 2020, doi: 10.1109/TGRS.2020.3001401.

[44] H. Liu, K. Simonyan, and Y. Yang, "DARTS: Differentiable architecture search," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2019, pp. 1–13.

[45] D. Marcos *et al.*, "Learning deep structure active contours end-to-end," in *Proc. Comput. Vision Pattern Recognit. (CVPR)*, 2018, pp. 8877–8885.

[46] Z. Li, J. D. Wegner, and A. Lucchi, "Topological map extraction from overhead images," in *Proc. IEEE Int. Conf. Comput. Vision*, 2019, pp. 1715–1724, doi: 10.1109/ICCV.2019.00180.

[47] N. Girard, D. Smirnov, J. Solomon, and Y. Tarabalka, "Polygonal building segmentation by frame field learning," Apr. 2020. [Online]. Available: http://arxiv.org/abs/2004.14875

[48] W. Zhao, C. Persello, and A. Stein, "Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework," *ISPRS J. Photogrammetry Remote Sens.*, vol. 175, pp. 119–131, May 2021, doi: 10.1016/j.isprsjprs.2021.02.014.

[49] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

[50] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6638947.

[51] O. Vinyals and Q. Le, "A neural conversational model," 2015, arXiv:1506.05869.

[52] M. Rußwurm and M. Körner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2017, pp. 1496–1504, doi: 10.1109/CVPRW.2017.193.

[53] M. Rußwurm, S. Lefèvre, and M. Körner, "Breizhcrops: A satellite time series dataset for crop type identification," in *Proc. ICML Workshop*, 2019, pp. 1–5.

[54] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, "Satellite image time series classification with pixel-set encoders and temporal self-attention," 2019, arXiv:1911.07757.

[55] M. O. Turkoglu, S. D'Aronco, J. D. Wegner, and K. Schindler, "Gating revisited: Deep multi-layer RNNs that can be trained," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, 2021, doi: 10.1109/TPAMI.2021.3064878.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[57] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop*, 2014.

[58] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learning*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/2200000006.

[59] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[60] T. Hank *et al.*, "Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges," *Surveys Geophys.*, vol. 40, no. 3, pp. 515–551, 2018, doi: 10.1007/s10712-018-9492-0.

[61] A. I. Flores-Anderson *et al.*, "Hyperspectral satellite remote sensing of water quality in Lake Atitlan, Guatemala," *Front. Environ. Sci.*, vol. 8, p. 7, 2020, doi: 10.3389/fenvs.2020.00007.

[62] J. P. Ryan, C. O. Davis, N. B. Tufillaro, R. M. Kudela, and B.-C. Gao, "Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in Monterey Bay, CA, USA," *Remote Sens.*, vol. 6, no. 2, pp. 1007–1025, 2014, doi: 10.3390/rs6021007.

[63] M. E. Fagan *et al.*, "Mapping species composition of forests and tree plantations in northeastern Costa Rica with an integration of hyperspectral and multitemporal Landsat imagery," *Remote Sens.*, vol. 7, no. 5, pp. 5660–5696, 2015, doi: 10.3390/rs70505660.

[64] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 5, no. 1, pp. 8–32, 2017, doi: 10.1109/MGRS.2016.2616418.

[65] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016, doi: 10.1109/TGRS.2016.2543748.

[66] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016, doi: 10.1109/TGRS.2016.2584107.

[67] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, 2018, doi: 10.1109/TGRS.2018.2794326.

[68] Y. Kong, X. Wang, and Y. Cheng, "Spectral-spatial feature extraction for HSI classification based on supervised hypergraph and sample expanded CNN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4128–4140, Nov. 2018, doi: 10.1109/JSTARS.2018.2869210.

[69] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3246–3263, 2020, doi: 10.1109/TGRS.2019.2951445.

[70] L. Mou, P. Ghamisi, and X. X. Zhu, "Unsupervised spectral–spatial feature learning via deep residual conv–deconv network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 391–406, 2018, doi: 10.1109/TGRS.2017.2748160.

[71] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral-spatial LSTMS," *Neurocomputing*, vol. 328, pp. 39–47, Feb. 2019, doi: 10.1016/j.neucom.2018.02.105.

[72] Q. Liu, F. Zhou, R. Hang, and X. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sens.*, vol. 9, no. 12, p. 1330, 2017, doi: 10.3390/rs9121330.

[73] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, 2019, doi: 10.1109/TGRS.2019.2899129.

[74] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral-spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5893–5909, 2018, doi: 10.1109/TGRS.2018.2827407.

[75] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, 2017, doi: 10.1109/TGRS.2016.2636241.

[76] X. Zhu *et al.*, "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, early access, 2021, doi: 10.1109/MGRS.2020.3046356.

[77] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, 2015, doi: 10.1109/LGRS.2015.2478256.

[78] Z. Zhang, H. Wang, F. Xu, and Y. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, 2017, doi: 10.1109/TGRS.2017.2743222.

[79] A. G. Mullissa, C. Persello, and A. Stein, "PolSARNET: A deep fully convolutional network for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, 2019, doi: 10.1109/JSTARS.2019.2956650.

[80] H. Dong, B. Zou, L. Zhang, and S. Zhang, "Automatic design of CNNs via differentiable neural architecture search for polSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6362–6375, 2020, doi: 10.1109/TGRS.2020.2976694.

[81] Z. Huang, M. Datcu, Z. Pan, and B. Lei, "Deep SAR-NET: Learning objects from signals," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 179–193, Jan. 2020, doi: 10.1016/j.isprsjprs.2020.01.016.

[82] S. Kazemi, B. Yonel, and B. Yazici, "Deep learning for direct automatic target recognition from SAR data," in *Proc. IEEE Radar Conf. (RadarConf)*, 2019, pp. 1–6, doi: 10.1109/RADAR.2019.8835492.

[83] K. Ichikawa and A. Hirose, "Singular unit restoration in INSAR using complex-valued neural networks in the spectral domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1717–1723, 2017, doi: 10.1109/TGRS.2016.2630719.

[84] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "ϕ-net: Deep residual learning for INSAR parameters estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 1–25, 2020, doi: 10.1109/TGRS.2020.3020427.

[85] F. Sica, F. Calvanese, G. Scarpa, and P. Rizzoli, "A CNN-based coherence-driven approach for INSAR phase unwrapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Oct. 2020, doi: 10.1109/LGRS.2020.3029565.

[86] A. Mazza, F. Sica, P. Rizzoli, and G. Scarpa, "Tandem-X forest mapping using convolutional neural networks," *Remote Sens.*, vol. 11, no. 24, p. 2980, Dec. 2019, doi: 10.3390/rs11242980.

[87] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, vol. 9351, pp. 234–241, Nov. 2015.

[88] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, 2017, doi: 10.1109/LSP.2017.2758203.

[89] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2017, pp. 5438–5441, doi: 10.1109/IGARSS.2017.8128234.

[90] A. G. Mullissa, C. Persello, and J. Reiche, "Despeckling polarimetric SAR data using a multistream complex-valued fully convolutional network," *IEEE Geosci. Remote Sens. Lett.*, early access, 2021, doi: 10.1109/LGRS.2021.3066311.

[91] T. Davis, V. Jain, A. Ley, O. D'Hondt, S. Valade, and O. Hellwich, "Reference-free despeckling of synthetic aperture radar images using a deep convolutional network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2020, pp. 3908–3911, doi: 10.1109/IGARSS39084.2020.9323293.

[92] X. Ma, C. Wang, Z. Yin, and P. Wu, "SAR image despeckling by noisy reference-based deep learning method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8807–8818, 2020, doi: 10.1109/TGRS.2020.2990978.

[93] P. Ghamisi *et al.*, "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 7, no. 1, pp. 6–39, 2019, doi: 10.1109/MGRS.2018.2890023.

[94] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Review: Deep learning on 3d point clouds," *Remote Sens.*, vol. 12, no. 11, p. 1729, May 2020, doi: 10.3390/rs12111729.

[95] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 1, no. 1, pp. 6–43, 2013, doi: 10.1109/MGRS.2013.2248301.

[96] S. Salcedo-Sanz *et al.*, "Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources," *Inf. Fusion*, vol. 63, pp. 256–272, Nov. 2020, doi: 10.1016/j.inffus.2020.07.004.

[97] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016, doi: 10.3390/rs8070594.

[98] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, 2018, doi: 10.1109/TGRS.2018.2817393.

[99] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 821–829, 2018, doi: 10.1109/JSTARS.2018.2797894.

[100] M. Belgiu and A. Stein, "Spatiotemporal image fusion in remote sensing," *Remote Sens.*, vol. 11, no. 7, p. 818, Apr. 2019, doi: 10.3390/rs11070818.

[101] L. H. Hughes, D. Marcos, S. Lobry, D. Tuia, and M. Schmitt, "A deep learning framework for matching of sar and optical imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 166–179, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.09.012.

[102] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, "Classification of hyperspectral and lidar data using coupled CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4939–4950, 2020, doi: 10.1109/TGRS.2020.2969024.

[103] J. Kang, M. Korner, Y. Wang, H. Taubenbock, and X. X. Zhu, "Building instance classification using street view images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 44–59, 2018, doi: 10.1016/j.isprsjprs.2018.02.006.

[104] C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 116, pp. 55–72, 2016, doi: 10.1016/j.isprsjprs.2016.03.008.

[105] W. F. Laurance, J. Sayer, and K. G. Cassman, "Agricultural expansion and its impacts on tropical nature," *Trends Ecol. Evolution*, vol. 29, no. 2, pp. 107–116, 2014, doi: 10.1016/j.tree.2013.12.001.

[106] F. Herzog, V. Prasuhn, E. Spiess, and W. Richner, "Environmental cross-compliance mitigates nitrogen and phosphorus pollution from Swiss agriculture," *Environ. Sci. Policy*, vol. 11, no. 7, pp. 655–668, 2008, doi: 10.1016/j.envsci.2008.06.003.

[107] N. B. Dise *et al.*, *Nitrogen as a Threat to European Terrestrial Biodiversity*. New York, NY, USA: Cambridge Univ. Press, 2011.

[108] S. K. Lowder, J. Skoet, and T. Raney, "The number, size, and distribution of farms, smallholder farms, and family farms worldwide," *World Develop.*, vol. 87, pp. 16–29, Nov. 2016, doi: 10.1016/j.worlddev.2015.10.041.

[109] R. Finger and N. Lehmann, "Policy reforms to promote efficient and sustainable water use in swiss agriculture," *Water Policy*, vol. 14, no. 5, pp. 887–901, 2012, doi: 10.2166/wp.2012.152.

[110] P. S. Thenkabail, I. Mariotto, M. K. Gumma, E. M. Middleton, D. R. Landis, and K. F. Huemmrich, "Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIS) for biophysical characterization and discrimination of crop types using field reflectance and hyperion/EO-1 data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 427–439, 2013, doi: 10.1109/JSTARS.2013.2252601.

[111] A. Walter, F. Liebisch, and A. Hund, "Plant phenotyping: from bean weighing to image analysis," *Plant Methods*, vol. 11, no. 1, pp. 1–11, 2015, doi: 10.1186/s13007-015-0056-8.

[112] J. Anderegg, K. Yu, H. Aasen, A. Walter, F. Liebisch, and A. Hund, "Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm," *Front. Plant Sci.*, vol. 10, p. 1749, Jan. 2020, doi: 10.3389/fpls.2019.01749.

[113] J. Inglada *et al.*, "Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery," *Remote Sens.*, vol. 7, no. 9, pp. 12,356–12,379, 2015, doi: 10.3390/rs70912356.

[114] F. Vuolo, M. Neuwirth, M. Immitzer, C. Atzberger, and W.-T. Ng, "How much does multi-temporal sentinel-2 data improve crop type classification?" *Int. J. Appl. Earth Observ. Geoinf.*, vol. 72, pp. 122–130, Oct. 2018, doi: 10.1016/j.jag.2018.06.007.

[115] S. Foerster, K. Kaden, M. Foerster, and S. Itzerott, "Crop type mapping using spectral–temporal profiles and phenological information," *Comput. Electron. Agriculture*, vol. 89, pp. 30–40, Nov. 2012, doi: 10.1016/j.compag.2012.07.015.

[116] M. Ustuner, F. B. Sanli, S. Abdikan, M. Esetlili, and Y. Kurucu, "Crop type classification using vegetation indices of rapideye imagery," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inform. Sci.*, vol. XL-7, no. 7, p. 195, 2014, doi: 10.5194/isprsarchives-XL-7-195-2014.

[117] C. Conrad *et al.*, "Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal rapideye images," *Comput. Electron. Agriculture*, vol. 103, pp. 63–74, Apr. 2014, doi: 10.1016/j.compag.2014.02.003.

[118] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "A hidden Markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data," *Remote Sens.*, vol. 7, no. 4, pp. 3633–3650, 2015, doi: 10.3390/rs70403633.

[119] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sens. Environ.*, vol. 204, pp. 509–523, Jan. 2018, doi: 10.1016/j.rse.2017.10.005.

[120] S. Bailly, S. Giordano, L. Landrieu, and N. Chehata, "Crop-rotation structured classification using multi-source sentinel images and LPIs for crop type mapping," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS 2018)*, 2018, pp. 1950–1953, doi: 10.1109/IGARSS.2018.8518427.

[121] M. Campos-Taberner *et al.*, "Understanding deep learning in land use classification based on sentinel-2 time series," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-74215-5.

[122] A. Rodriguez and J. Wegner, "Counting the uncountable: Deep semantic density estimation from space," in *Proc. German Conf. Pattern Recognit., LNCS 11269*, 2018, pp. 351–362.

[123] M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, p. 129, 2018, doi: 10.3390/ijgi7040129.

[124] R. Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, "Semantic segmentation of crop type in Africa: A novel dataset and analysis of deep learning methods," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2019, pp. 76–82.

[125] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, pp. 523, 2019, doi: 10.3390/rs11050523.

[126] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[127] N. Metzger, M. O. Turkoglu, S. D'Aronco, J. D. Wegner, and K. Schindler, "Crop classification under varying cloud cover with neural ordinary differential equations," 2020, arXiv 2012.02542.

[128] A. Rydberg and G. Borgefors, "Integrated method for boundary delineation of agricultural fields in multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2514–2520, 2001, doi: 10.1109/36.964989.

[129] L. Yan and D. P. Roy, "Automated crop field extraction from multi-temporal Web Enabled Landsat Data," *Remote Sens. Environ.*, vol. 144, pp. 42–64, Mar. 2014, doi: 10.1016/j.rse.2014.01.006.

[130] J. Graesser and N. Ramankutty, "Detection of cropland field parcels from Landsat imagery," *Remote Sens. Environ.*, vol. 201, pp. 165–180, Aug. 2017, doi: 10.1016/j.rse.2017.08.027.

[131] M. Lesiv *et al.*, "Estimating the global distribution of field size using crowdsourcing," *Global Change Biol.*, vol. 25, no. 1, pp. 174–186, Jan. 2019, doi: 10.1111/gcb.14492.

[132] C. Persello, V. A. Tolpekin, J. R. Bergado, and R. A. de By, "Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping," *Remote Sens. Environ.*, vol. 231, p. 111253, Sep. 2019, doi: 10.1016/j.rse.2019.111253.

[133] S. Marvaniya, U. Devi, J. Hazra, S. Mujumdar, and N. Gupta, "Small, sparse, but substantial: Techniques for segmenting small agricultural fields using sparse ground data," *Int. J. Remote Sens.*, vol. 42, no. 4, pp. 1512–1534, 2020, doi: 10.1080/01431161.2020.1834166.

[134] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 1–14, 2017, doi: 10.1109/TPAMI.2016.2644615.

[135] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2011, doi: 10.1109/TPAMI.2010.161.

[136] J. Pont-Tuset, P. Arbelaez, J. T. B. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping for image segmentation and object proposal generation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 128–140, 2017, doi: 10.1109/TPAMI.2016.2537320.

[137] A. Garcia-Pedrero, M. Lillo-Saavedra, D. Rodriguez-Esparragon, and C. Gonzalo-Martin, "Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system," *IEEE Access*, vol. 7, pp. 158,223–158,236, Oct. 2019, doi: 10.1109/ACCESS.2019.2950371.

[138] F. Waldner and F. I. Diakogiannis, "Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network," *Remote Sens. Environ.*, vol. 245, p. 111741, Aug. 2020, doi: 10.1016/j.rse.2020.111741.

[139] K. M. Masoud, C. Persello, and V. A. Tolpekin, "Delineation of agricultural field boundaries from sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks," *Remote Sens.*, vol. 12, no. 1, p. 59, Dec. 2019, doi: 10.3390/rs12010059.

[140] M. P. Wagner and N. Oppelt, "Deep learning and adaptive graph-based growing contours for agricultural field extraction," *Remote Sens.*, vol. 12, no. 12, p. 1990, 2020, doi: 10.3390/rs12121990.

[141] "World urbanization prospects: The 2018 revision," United Nations, Tech. Rep., 2018. [Online]. Available: https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf

[142] "Habitat III issue papers 22-informal settlements," UN-Habitat, 2015. [Online]. Available: https://unhabitat.org/sites/default/files/download-manager-files/Habitat-III-Issue-Paper-22_Informal-Settlements-2.0%20%282%29.pdf

[143] "Addressing the most vulnerable first pro-poor climate action in informal settlements," UN-Habitat, 2018. [Online]. Available: https://unhabitat.org/pro-poor-climate-action-in-informal-settlement

[144] "Slum almanac 2015 2016," UN-Habitat, 2015. [Online]. Available: https://unhabitat.org/slum-almanac-2015-2016-0

[145] M. Kuffer et al., "The scope of earth-observation to improve the consistency of the SDG slum indicator," ISPRS Int. J. Geo-Inf., vol. 7, no. 11, p. 428, Nov. 2018, doi: 10.3390/ijgi7110428.

[146] M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from Space—15 years of slum mapping using remote sensing," Remote Sens., vol. 8, no. 6, p. 455, May 2016, doi: 10.3390/rs8060455.

[147] C. M. Gevaert, C. Persello, R. Sliuzas, and G. Vosselman, "Informal settlement classification using point-cloud and image-based features from UAV data," ISPRS J. Photogrammetry Remote Sens., vol. 125, pp. 225–236, Mar. 2017, doi: 10.1016/j.isprsjprs.2017.01.017.

[148] C. M. Gevaert, C. Persello, R. Sliuzas, and G. Vosselman, "Monitoring household upgrading in unplanned settlements with unmanned aerial vehicles," Int. J. Appl. Earth Observ. Geoinf., vol. 90, p. 102,117, Aug. 2020.

[149] J. Graesser, A. Cheriyadat, R. R. Vatsavai, V. Chandola, J. Long, and E. Bright, "Image based characterization of formal and informal neighborhoods in an urban landscape," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 5, no. 4, pp. 1164–1176, 2012, doi: 10.1109/JSTARS.2012.2190383.

[150] M. Kuffer, K. Pfeffer, R. Sliuzas, and I. Baud, "Extraction of slum areas from VHR imagery using GLCM variance," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 5, pp. 1830–1840, 2016, doi: 10.1109/JSTARS.2016.2538563.

[151] N. Mboga, C. Persello, J. Bergado, and A. Stein, "Detection of informal settlements from VHR images using convolutional neural networks," Remote Sens., vol. 9, no. 11, p. 1106, Oct. 2017, doi: 10.3390/rs9111106.

[152] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenbock, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," ISPRS J. Photogrammetry Remote Sens., vol. 150, pp. 59–69, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.006.

[153] J. Wang, M. Kuffer, D. Roy, and K. Pfeffer, "Deprivation pockets through the lens of convolutional neural networks," Remote Sens. Environ., vol. 234, p. 111448, Dec. 2019, doi: 10.1016/j.rse.2019.111448.

[154] R. Liu, M. Kuffer, and C. Persello, "The temporal dynamics of slums employing a CNN-based change detection approach," Remote Sens., vol. 11, no. 23, pp. 2844, 11 2019, doi: 10.3390/rs11232844.

[155] A. Ajami, M. Kuffer, C. Persello, and K. Pfeffer, "Identifying a slums' degree of deprivation from VHR images using convolu-

[156] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," Science, vol. 353, no. 6301, pp. 790–794, 2016, doi: 10.1126/science.aaf7894.

[157] E. Suel, J. W. Polak, J. E. Bennett, and M. Ezzati, "Measuring social, environmental and health inequalities using deep learning and street imagery," Sci. Rep., vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-42036-w.

[158] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Machine Learning, 2019, pp. 6105–6114.

[159] J. L. Abitbol and M. Karsai, "Socioeconomic correlations of urban patterns inferred from aerial images: Interpreting activation maps of convolutional neural networks," Nature Mach. Intell., vol. 2, no. 11, pp. 684–692, 2020, doi: 10.1038/s42256-020-00243-5.

[160] P. Dale and J. McLaughlin, Land Administration. London, U.K.: Oxford Univ. Press, 1999.

[161] S. Enemark, K. Bell, C. Lemmen, and R. McLaren, Fit-for-purpose Land Administration, Joint Fig ed., The World Bank and the International Federation of Surveyors (FIG), Copenhagen, Denmark, 2014.

[162] C. Boone, "Legal empowerment of the poor through property rights reform: Tensions and trade-offs of land registration and titling in Sub-Saharan Africa," J. Develop. Stud., vol. 55, no. 3, pp. 384–400, Mar. 2019, doi: 10.1080/00220388.2018.1451633.

[163] C. Lakner, D. G. Mahler, M. Negre, and E. B. Prydz, "How much does reducing inequality matter for global poverty?" World Bank Group, Tech. Rep., 2019. [Online]. Available: https://openknowledge.worldbank.org/bitstream/handle/10986/33902/How-Much-Does-Reducing-Inequality-Matter-for-Global-Poverty.pdf?sequence=1&isAllowed=y

[164] P. F. Conforti, Looking Ahead in World Food and Agriculture: Perspectives to 2050. 2011. [Online]. Available: https://www.fao.org/3/i2280e/i2280e.pdf

[165] L. Alden Wily, "Collective land ownership in the 21st century: Overview of global trends," Land, vol. 7, no. 2, p. 68, May 2018, doi: 10.3390/land7020068.

[166] R. Meinzen-Dick, A. Quisumbing, C. Doss, and S. Theis, "Women's land rights as a pathway to poverty reduction: Framework and review of available evidence," Agricultural Syst., vol. 172, pp. 72–82, Jun. 2019, doi: 10.1016/j.agsy.2017.10.009.

[167] I. Williamson, "The justification of cadastral systems in developing countries," Geomatica, vol. 51, no. 1, pp. 21–36, 1997.

[168] M. Koeva et al., "Innovative geospatial solutions for land tenure mapping," Rwanda J. Eng., Sci., Technol. Environ., vol. 3, no. 1, pp. 34–49, Jul. 2020, doi: 10.4314/rjeste.v3i1.3S.

[169] S. Enemark, R. McLaren, C. Lemmen, D. Antonio, and J. Gitau, Guiding Principles for Building fit-for-purpose Land Administration Systems in Developing Countries. Nairobi: UN-Habitat, 2016.

[170] "Framework for effective land administration: A reference for developing, reforming, renewing, strengthening or modernizing land administration and management systems,"

Expert Group on Land Administration and Management, United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), 2019. [Online]. Available: https://ggim.un.org/meetings/GGIM -committee/9th-Session/documents/E_C.20_2020_10_ Add_1_LAM_background.pdf

[171] I. Williamson, S. Enemark, J. Wallace, and A. Rajabifard, *Land Administration for Sustainable Development*. Redlands: ESRI Press Academics, 2010.

[172] X. Luo, R. Bennett, M. Koeva, C. Lemmen, and N. Quadros, "Quantifying the overlap between cadastral and visual boundaries: A case study from Vanuatu," *Urban Sci.*, vol. 1, no. 4, p. 32, Nov. 2017, doi: 10.3390/urbansci1040032.

[173] S. Crommelinck, R. Bennett, M. Gerke, F. Nex, M. Y. Yang, and G. Vosselman, "Review of automatic feature extraction from high-resolution optical sensor data for UAV-based cadastral mapping," *Remote Sens.*, vol. 8, no. 8, p. 689, 2016, doi: 10.3390/ rs8080689.

[174] M. Manyoky, P. Theiler, D. Steudler, and H. Eisenbeiss, "Unmanned aerial vehicle in cadastral applications," *ISPRS – Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XXX-VIII-1/C22, pp. 57–62, Sep. 2012, doi: 10.5194/isprsarchives-XXXVIII-1-C22-57-2011.

[175] I. Jazayeri, A. Rajabifard, and M. Kalantari, "A geometric and semantic evaluation of 3D data sourcing methods for land and property information," *Land Use Policy*, vol. 36, pp. 219–230, Jan. 2014, doi: 10.1016/j.landusepol.2013.08.004.

[176] M. Koeva, M. Muneza, C. Gevaert, M. Gerke, and F. Nex, "Using UAVs for map creation and updating. A case study in Rwanda," *Survey Rev.*, vol. 50, no. 361, pp. 1–14, 2016, doi: 10.1080/00396265.2016.1268756.

[177] A. García-Pedrero, C. Gonzalo-Martín, and M. Lillo-Saavedra, "A machine learning approach for agricultural parcel delineation through agglomerative segmentation," *Int. J. Remote Sens.*, vol. 38, no. 7, pp. 1809–1819, Apr. 2017, doi: 10.1080/01431161.2016.1278312.

[178] O. E. Dragomir, R. Gouriveau, F. Dragomir, E. Minca, and N. Zerhouni, "Review of prognostic problem in condition-based maintenance," in *Proc. European Control Conf. (ECC)*, 2009, pp. 1587–1592, doi: 10.23919/ECC.2009.7074633.

[179] S. Crommelinck, M. Koeva, M. Y. Yang, and G. Vosselman, "Application of deep learning for delineation of visible cadastral boundaries from remote sensing imagery," *Remote Sens.*, vol. 11, no. 21, pp. 1–22, 2019, doi: 10.3390/rs11212505.

[180] X. Xia, C. Persello, and M. Koeva, "Deep fully convolutional networks for cadastral boundary detection from UAV images," *Remote Sens.*, vol. 11, no. 14, p. 1725, Jul. 2019, doi: 10.3390/rs11141725.

[181] M. Koeva, S. Crommelinck, C. Stöcker, and J. Crompvoets, "Its4land – Challenges and opportunities in developing innovative geospatial tools for fit-for-purpose land rights mapping," in *Proc. FIG Congr.*, May 2018, pp. 1–17.

[182] M. Koeva *et al.*, "Innovative remote sensing methodologies for Kenyan land tenure mapping," *Remote Sens.*, vol. 12, no. 2, pp. 1–27, Jan. 2020, doi: 10.3390/rs12020273.

[183] M. N. Koeva *et al.*, "Towards innovative geospatial tools for fit-for-purpose land rights mapping," *Int. Arch. Photogrammetry, Re-mote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 37–43, Sep. 2017, doi: 10.5194/isprs-archives-XLII-2-W7-37-2017.

[184] S. Crommelinck, B. Höfle, M. Koeva, M. Yang, and G. Vosselman, "Interactive cadastral boundary delineation from UAV data," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2, no. 2, pp. 81–88, 2018, doi: 10.5194/isprs-annals-IV-2-81-2018.

[185] M. Chipofya, M. Karamesouti, C. Schultz, and A. Schwering, "Local domain models for land tenure documentation and their interpretation into the LADM," *Land Use Policy*, vol. 99, p. 105,005, Dec. 2020, doi: 10.1016/j.landusepol.2020.105005.

[186] J. Romm, *Climate Change: What Everyone Needs to Know*. London, U.K.: Oxford Univ. Press, 2018.

[187] D. Archer and S. Rahmstorf, *The Climate Crisis: An Introductory Guide to Climate Change*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[188] C. B. Field, V. Barros, T. F. Stocker, and Q. Dahe, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[189] D. Rolnick *et al.*, "Tackling climate change with machine learning," 2019, arXiv:1906.05433.

[190] J. Mathe, N. Miolane, N. Sebastien, and J. Lequeux, "PVNet: A LRCN architecture for spatio-temporal photovoltaic power forecasting from numerical weather prediction," 2020, arXiv:1902.01453.

[191] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012, doi: 10.1016/j.renene.2011.05.033.

[192] C. Elkin and S. Witherspoon, "Machine learning can boost the value of wind energy," DeepMind, Feb. 26, 2019. [Online]. Available: https://deepmind.com/blog/machine-learning-can-boost-value-of-wind-energy/

[193] P. Donti, B. Amos, and J. Z. Kolter, "Task-based end-to-end model learning in stochastic optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5484–5494.

[194] J. M. Malof, K. Bradbury, L. M. Collins, and R. G. Newell, "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery," *Appl. Energy*, vol. 183, pp. 229–240, Dec. 2016, doi: 10.1016/j.apenergy.2016.08.191.

[195] J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal, "DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States," *Joule*, vol. 2, no. 12, pp. 2605–2617, 2018, doi: 10.1016/j.joule.2018.11.021.

[196] H. Jiang and Y. Zhang, "Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2016, pp. 1–5, doi: 10.1109/ PESGM.2016.7741933.

[197] M. Pertl, K. Heussen, O. Gehrke, and M. Rezkalla, "Voltage estimation in active distribution grids using neural networks," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, 2016, pp. 1–5, doi: 10.1109/PESGM.2016.7741758.

[198] B. Bhattacharya and A. Sinha, "Deep fault analysis and subset selection in solar power grids," 2017, arXiv:1711.02810v1.

[199] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 107–120, Jul. 2018, doi: 10.1016/j.ijepes.2017.12.016.

[200] Q. Jiang, L. Cao, M. Cheng, C. Wang, and J. Li, "Deep neural networks-based vehicle detection in satellite images," in *Proc. Int. Symp. Bioelectron. Bioinform. (ISBB)*, 2015, pp. 184–187, doi: 10.1109/ISBB.2015.7344954.

[201] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, "A large contextual dataset for classification detection and counting of cars with deep learning," in *Proc. European Conf. Comput. Vision*, 2016, pp. 785–800.

[202] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3562–3664, 2017, doi: 10.1109/JSTARS.2017.2694890.

[203] L. H. Kaack, G. H. Chen, and M. G. Morgan, "Truck traffic monitoring with satellite images," in *Proc. 2nd ACM SIGCAS Conf. Comput. Sustain. Soc.*, 2019, pp. 155–164.

[204] F. Jia, Y. Lei, J. Lin, X. Zhou, and A. Na Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016, doi: 10.1016/j.ymssp.2015.10.025.

[205] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning," *Energy Buildings*, vol. 177, pp. 12–22, Oct. 2018, doi: 10.1016/j.enbuild.2018.08.010.

[206] M. Ateeq, F. Ishmanov, M. K. Afzal, and M. Naeem, "Multi-parametric analysis of reliability and energy consumption in IoT: A deep learning approach," *Sensors*, vol. 19, no. 2, p. 2, 2019, doi: 10.3390/s19020309.

[207] S. Mo, Y. Zhu, N. Zabaras, X. Shi, and J. Wu, "Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media," *Water Resour. Res.*, vol. 55, no. 1, pp. 703–728, 2019, doi: 10.1029/2018WR023528.

[208] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, "Could machine learning break the convection parameterization deadlock?" *Geophys. Res. Lett.*, vol. 45, no. 11, pp. 5742–5751, 2018, doi: 10.1029/2018GL078202.

[209] G. Camps-Valls, D. Tuia, X. X. Zhu, and M. R. (Editors), *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*. Hoboken, NJ, USA: Wiley, 2021.

[210] J. Bolibar, A. Rabatel, I. Gouttevin, C. Galiez, T. Condom, and E. Sauquet, "Deep learning applied to glacier evolution modelling," *Cryosphere*, vol. 14, no. 2, pp. 565–584, 2020, doi: 10.5194/tc-14-565-2020.

[211] V. Nieves, C. Radin, and G. Camps-Valls, "Predicting regional coastal sea level changes with machine learning," *Sci. Rep.*, 2021, to be published.

[212] W. Li, L. Ni, Z. liang Li, S. bo Duan, and H. Wu, "Evaluation of machine learning algorithms in spatial downscaling of MODIS land surface temperature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2299–2307, 2019, doi: 10.1109/JSTARS.2019.2896923.

[213] M. C. Perignon, P. Passalacqua, T. M. Jarriel, J. M. Adams, and I. Overeem, "Patterns of geomorphic processes across deltas using image analysis and machine learning," in *AGU Fall Meeting Abstracts*, Washington, D.C.: American Geophysical Union, 2018.

[214] L. Bragilevsky and I. V. Bajić, "Deep learning for Amazon satellite image analysis," in *Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM)*, 2017, pp. 1–5, doi: 10.1109/PACRIM.2017.8121895.

[215] A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in *Proc. 1st ACM SIGCAS Conf. Comput. Sustain. Soc.*, 2018, p. 50.

[216] S. Voigt, T. Kemper, T. Riedlinger, R. Kiefl, K. Scholte, and H. Mehl, "Satellite image analysis for disaster and crisis-management support," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1520–1528, 2007, doi: 10.1109/TGRS.2007.895830.

[217] C. S. de Witt and T. Hornigold, "Stratospheric aerosol injection as a deep reinforcement learning problem," 2019, arXiv:1905.07366.

[218] Q. Di, I. Kloog, P. Koutrakis, A. Lyapustin, Y. Wang, and J. Schwartz, "Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States," *Environ. Sci. Technol.*, vol. 50, no. 9, pp. 4712–4721, 2016, doi: 10.1021/acs.est.5b06121.

[219] A. Crane-Droesch, B. Kravitz, and J. T. Abatzoglou, "Using deep learning to model potential impacts of geoengineering via solar radiation management on US agriculture," in *AGU Fall Meeting Abstracts*, Washington, D.C.: American Geophysical Union, 2018.

[220] "The state of the world's forests 2018 – Forest pathways to sustainable development," Food and Agriculture Organization. http://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1144279/ (accessed Dec. 5, 2020).

[221] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks," *Remote Sens.*, vol. 11, no. 11, p. 1309, Jun. 2019, doi: 10.3390/rs11111309.

[222] T. Dong, Y. Shen, J. Zhang, Y. Ye, and J. Fan, "Progressive cascaded convolutional neural networks for single tree detection with google earth imagery," *Remote Sens.*, vol. 11, no. 15, p. 1786, Jul. 2019, doi: 10.3390/rs11151786.

[223] Z. M. Hamdi, M. Brandmeier, and C. Straub, "Forest damage assessment using deep learning on high resolution remote sensing data," *Remote Sens.*, vol. 11, no. 17, p. 1976, Aug. 2019, doi: 10.3390/rs11171976.

[224] "Planet: Understanding the amazon from space," Kaggle, 2017. https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/ (accessed Dec. 5, 2020).

[225] S.-H. Lee, K.-J. Han, K. Lee, K.-J. Lee, K.-Y. Oh, and M.-J. Lee, "Classification of landscape affected by deforestation using high-resolution remote sensing data and deep-learning techniques," *Remote Sens.*, vol. 12, no. 20, p. 3372, Oct. 2020, doi: 10.3390/rs12203372.

[226] P. de Bem, O. de Carvalho Junior, R. Fontes Guimarães, and R. Trancoso Gomes, "Change detection of deforestation in the

Brazilian Amazon using Landsat data and convolutional neural networks," *Remote Sens.*, vol. 12, no. 6, p. 901, Mar. 2020, doi: 10.3390/rs12060901.

[227] L. Bragilevsky and I. V. Bajić, "Deep learning for amazon satellite image analysis," in *Proc. IEEE Pacific Rim Conf. Commun., Comput. Signal Process. (PACRIM)*, 2017, pp. 1–5, doi: 10.1109/PACRIM.2017.8121895.

[228] R. V. Maretto, L. M. G. Fonseca, N. Jacobs, T. S. Körting, H. N. Bendini, and L. L. Parente, "Spatio-temporal deep learning approach to map deforestation in amazon rainforest," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 1–5, 2020, doi: 10.1109/LGRS.2020.2986407.

[229] J. Irvin *et al.*, "Forestnet: Classifying drivers of deforestation in Indonesia using deep learning on satellite imagery," in *Proc. NeurIPS 2020 Workshop Tackling Climate Change Machine Learning*, Dec. 2020, pp. 1–10.

[230] K. Govil, M. L. Welch, J. T. Ball, and C. R. Pennypacker, "Preliminary results from a wildfire detection system using deep learning on remote camera images," *Remote Sens.*, vol. 12, no. 1, p. 166, Jan 2020, doi: 10.3390/rs12010166.

[231] Y. Zhao, J. Ma, X. Li, and J. Zhang, "Saliency detection and deep learning-based wildfire identification in UAV imagery," *Sensors*, vol. 18, no. 3, p. 712, Feb. 2018, doi: 10.3390/s18030712.

[232] N. T. Toan, P. T Cong, N. Q. Viet Hung, and J. Jo, "A deep learning approach for early wildfire detection from hyperspectral satellite images," in *Proc. 7th Int. Conf. Robot Intell. Technol. Appl. (RiTA)*, 2019, pp. 38–45, doi: 10.1109/RITAPP.2019.8932740.

[233] Y. Ban, P. Zhang, A. Nascetti, A. R. Bevington, and M. A. Wulder, "Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning," *Sci. Rep.*, vol. 10, no. 1, p. 1322, Jan. 2020, doi: 10.1038/s41598-019-56967-x.

[234] D. Radke, A. Hessler, and D. Ellsworth, "Firecast: Leveraging deep learning to predict wildfire spread," in *Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI-19)*, Jul. 2019, pp. 4575–4581.

[235] J. R. Bergado, C. Persello, K. Reinke, and A. Stein, "Predicting wildfire burns from big geodata using deep learning," *Safety Sci.*, vol. 140, p. 105,276, Aug. 2021, doi: 10.1016/j.ssci.2021.105276.

[236] A. Farasin, L. Colomba, and P. Garza, "Double-step U-Net: A deep learning-based approach for the estimation of wildfire damage severity through Sentinel-2 satellite data," *Appl. Sci.*, vol. 10, no. 12, p. 4332, Jun. 2020, doi: 10.3390/app10124332.

[237] W. Zhou and L. Klein, "Monitoring the impact of wildfires on tree species with deep learning," in *Proc. NeurIPS Workshop Tackling Climate Change with Machine Learning*, Dec. 2020, pp. 1–6.

[238] G. Camps-Valls *et al.*, "A unified vegetation index for quantifying the terrestrial biosphere," *Sci. Adv.*, vol. 7, no. 9, p. eabc7447, 2021, doi: 10.1126/sciadv.abc7447.

[239] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.

[240] M. W. Scown, "The sustainable development goals need geoscience," *Nature GeoSci.*, vol. 13, no. 11, pp. 714–715, Oct. 2020, doi: 10.1038/s41561-020-00652-6.

[241] J. Tan *et al.*, "Deep learning convolutional neural network for the retrieval of land surface temperature from AMSR2 data in China," *Sensors*, vol. 19, no. 13, p. 2987, 2019, doi: 10.3390/s19132987.

[242] J. Verrelst *et al.*, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 108, pp. 273–290, Oct. 2015, doi: 10.1016/j.isprsjprs.2015.05.005.

[243] F. J. García-Haro *et al.*, "A global canopy water content product from AVHRR/Metop," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, no. 7, pp. 1–18, 2020, doi: 10.1016/j.isprsjprs.2020.02.007.

[244] A. Moreno-Martinez *et al.*, "A methodology to derive global maps of leaf traits using remote sensing and climate data," *Remote Sens. Environ.*, vol. 218, no. 12, pp. 69–88, 2018, doi: 10.1016/j.rse.2018.09.006.

[245] M. Jung *et al.*, "The FLUXCOM ensemble of global land-atmosphere energy fluxes," *Sci. Data*, vol. 6, no. 1, 2019, doi: 10.1038/s41597-019-0076-8.

[246] M. Jung *et al.*, "Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach," *Biogeosciences*, vol. 17, no. 5, pp. 1343–1365, 2020, doi: 10.5194/bg-17-1343-2020.

[247] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agriculture*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[248] K. Kuwata and R. Shibasaki, "Estimating corn yield in the United States with Modis Evi and machine learning methods," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inform. Sci.*, vol. 3, no. 8, 2016.

[249] A. Wolanin *et al.*, "Estimating and understanding crop yields with explainable deep learning in the Indian wheat belt," *Environ. Res. Lett.*, vol. 15, no. 2, pp. 1–12, 2020, doi: 10.1088/1748-9326/ab68ac.

[250] S. K. Kim, S. Ames, J. Lee, C. Zhang, A. C. Wilson, and D. Williams, "Massive scale deep learning for detecting extreme climate events," in *Proc. 7th Int. Workshop Climate Informatics: CI 2017*, pp. 17–20.

[251] L. Ye, L. Gao, R. Marcos-Martinez, D. Mallants, and B. A. Bryan, "Projecting Australia's forest cover dynamics and exploring influential factors using deep learning," *Environ. Model. Softw.*, vol. 119, pp. 407–417, Sep. 2019, doi: 10.1016/j.envsoft.2019.07.013.

[252] S. H. Khan, X. He, F. Porikli, and M. Bennamoun, "Forest change detection in incomplete satellite images with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5407–5423, 2017, doi: 10.1109/TGRS.2017.2707528.

[253] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined lidar and Landsat 8 data," *Remote Sens.*, vol. 11, no. 12, p. 1459, 2019, doi: 10.3390/rs11121459.

[254] N. Lang, K. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sens. Environ.*, vol. 233, p. 111,347, Nov. 2019, doi: 10.1016/j.rse.2019.111347.

[255] G. Ceballos, P. R. Ehrlich, and P. H. Raven, "Vertebrates on the brink as indicators of biological annihilation and the

sixth mass extinction," *Proc. Natl. Acad. Sci.*, vol. 117, no. 24, pp. 13,596–13,602, 2020, doi: 10.1073/pnas.1922686117.

[256] H. S. Young, D. J. McCauley, M. Galetti, and R. Dirzo, "Patterns, causes, and consequences of anthropocene defaunation," *Annu. Rev. Ecol. Evol. Syst.*, vol. 47, no. 1, pp. 333–358, 2016, doi: 10.1146/annurev-ecolsys-112414-054142.

[257] I. Duporge, O. Isupova, S. Reece, D. W. Macdonald, and T. Wang, "Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes," *Remote Sens. Ecol. Conservation*, vol. 7, no. 3, pp. 369–381, 2020.

[258] N. Rey, M. Volpi, S. Joost, and D. Tuia, "Detecting animals in African Savanna with UAVs and the crowds," *Remote Sens. Environ.*, vol. 200, pp. 341–351, Oct. 2017, doi: 10.1016/j.rse.2017.08.026.

[259] J. Linchant, J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen, "Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges," *Mammal Rev.*, vol. 45, no. 4, pp. 239–252, 2015, doi: 10.1111/mam.12046.

[260] P. Fust and J. Loos, "Development perspectives for the application of autonomous, unmanned aerial systems (UASs) in wildlife conservation," *J. Biol. Conserv*, vol. 241, p. 108380, Jan. 2020, doi: 10.1016/j.biocon.2019.108380.

[261] B. Kellenberger, D. Marcos, and D. Tuia, "Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning," *Remote Sens. Environ.*, vol. 216, pp. 139–153, Oct. 2018, doi: 10.1016/j.rse.2018.06.028.

[262] M. S. Norouzzadeh *et al.*, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proc. Nat. Acad. Sci.*, vol. 115, no. 25, pp. E5716–E5725, 2018, doi: 10.1073/pnas.1719367115.

[263] J. A. J. Eikelboom *et al.*, "Improving the precision and accuracy of animal population estimates with aerial image object detection," *Methods Ecol. Evol.*, vol. 10, no. 11, pp. 1875–1887, 2019, doi: 10.1111/2041-210X.13277.

[264] J. Peng *et al.*, "Wild animal survey using UAS imagery and deep learning: modified faster R-CNN for Kiang detection in Tibetan plateau," *ISPRS J. Photo. Remote Sens.*, vol. 169, pp. 364–376, Nov. 2020, doi: 10.1016/j.isprsjprs.2020.08.026.

[265] S. Beery, D. Morris, and S. Yang, "Efficient pipeline for camera trap image review," 2019, arXiv:1907.06772.

[266] B. Kellenberger, D. Tuia, and D. Morris, "AIDE: Accelerating image-based ecological surveys with artificial intelligence," *Methods Ecol. Evol.*, vol. 11, no. 12, pp. 1716–1727, 2020, doi: 10.1111/2041-210X.13489.

[267] L. D. Brown and D. Schormann, "Poacher detection and wildlife counting system," in *Proc. Southern Africa Telecommun. Netw. Appl. Conf. (SATNAC)*, 2019, pp. 1–6.

[268] Y. Wang, Z. R. Shi, L. Yu, R. Singh, L. Joppa, and F. Fang, "Deep reinforcement learning for green security games with real-time information," in *Proc. Assoc. Adv. Artif. Intell.*, 2019, doi: 10.1609/aaai.v33i01.33011401.

[269] C. J. Torney *et al.*, "A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images," *Meth. Ecol. Evol.*, vol. 10, no. 6, pp. 779–787, 2019, doi: 10.1111/2041-210X.13165.

[270] B. Settles, *Active Learning.* San Rafael, CA, USA: Morgan and Claypool, 2012, doi: 10.2200/S00429ED1V01Y201207AIM018.

[271] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9524–9533, 2019, doi: 10.1109/TGRS.2019.2927393.

[272] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2017. pp. 5574–5584.

[273] S. Chakraborty *et al.*, "Interpretability of deep learning models: A survey of results," in *Proc. IEEE SmartWorld Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet of People Smart City Innov., SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI 2017*, 2018, pp. 1–6, doi: 10.1109/UIC-ATC.2017.8397411.

[274] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. R. Müller, "Toward interpretable machine learning: Transparent deep neural networks and beyond," 2020, arXiv:2003.07631.

[275] A. Levering, D. Marcos, and D. Tuia, "On the relation between landscape beauty and land cover: A case study in the U.K. at Sentinel-2 resolution with interpretable AI," *ISPRS J. Int. Soc. Photo. Remote Sens.*, vol. 177, pp. 194–203, Jul. 2021, doi: 10.1016/j.isprsjprs.2021.04.020.

[276] D. Marcos, S. Lobry, R. Fong, N. Courty, R. Flamary, and D. Tuia, "Contextual semantic interpretability," in *Proc. Asian Conf. Comput. Vision (ACCV)*, Kyoto, Kapan, 2020.

[277] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag. (replaces Newslett.)*, vol. 4, no. 2, pp. 41–57, 2016, doi: 10.1109/MGRS.2016.2548504.

[278] C. Doldirina, "Open data and earth observations the case of opening up access to and use of earth observation data through the global earth observation system of systems," *JIPITEC*, vol. 6, no. 1, pp. 73–85, 2015.

[279] S. Stall *et al.*, "Make scientific data fair," *Nature*, vol. 570, no. 7759, pp. 27–29, 2019, doi: 10.1038/d41586-019-01720-7.

[280] S. David, "*Perspectives on Free and Open Source Software*. Joseph Feller, Brian Fitzgerald, Scott A. Hissam, Karim R. Lakhani (Eds.)," *J. Inform. Technol. Politics*, vol. 4, no. 3, pp. 97–99, 2008, doi: 10.1080/19331680801916908.

[281] G. Boulton, "The challenges of a big data earth," *Big Earth Data*, vol. 2, no. 1, pp. 1–7, 2018, doi: 10.1080/20964471.2017.1397411.

[282] C. E. Woodcock *et al.*, "Free access to Landsat imagery," *Science*, vol. 320, no. 5879, p. 1011, 2008, doi: 10.1126/science.320.5879.1011a.

[283] T. Dhu *et al.*, "Digital earth Australia – Unlocking new value from earth observation data," *Big Earth Data*, vol. 1, nos. 1–2, pp. 64–74, 2017, doi: 10.1080/20964471.2017.1402490.

[284] M. Mahecha *et al.*, "Earth system data cubes unravel global multivariate dynamics," *Earth Syst. Dynam.*, vol. 11, no. 1, pp. 201–234, Feb. 2020, doi: 10.5194/esd-11-201-2020.

[285] 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, 2017, doi: 10.1080/20964471.2017.1398903.

GRS