Position: Mission Critical – Satellite Data is a Distinct Modality in Machine Learning

Esther Rolf [*1 2] **Konstantin Klemmer** [3] **Caleb Robinson** [4] **Hannah Kerner** [*5]

Abstract

Satellite data has the potential to inspire a seismic shift for machine learning—one in which we rethink existing practices designed for traditional data modalities. As machine learning for satellite data (SatML) gains traction for its real-world impact, our field is at a crossroads. We can either continue applying ill-suited approaches, or we can initiate a new research agenda that centers around the unique characteristics and challenges of satellite data. This position paper argues that satellite data constitutes a distinct modality for machine learning research and that we must recognize it as such to advance the quality and impact of SatML research across theory, methods, and deployment. We outline critical discussion questions and actionable suggestions to transform SatML from merely an intriguing application area to a dedicated research discipline that helps move the needle on big challenges for machine learning and society.

1. Introduction

Hundreds of remote sensing satellites continuously monitor the Earth's surface, creating petabyte-scale time series datasets. These datasets hold information needed to measure and address pressing planetary-scale challenges including climate change (Yang et al., 2013), poverty (Jean et al., 2016), food insecurity (Nakalembe, 2020), biodiversity loss (Skidmore et al., 2021), and many other global goals (Kavvada et al., 2020). However, extracting actionable insights from satellite data at scale requires specialized analysis.

[*]Equal contribution [1]Harvard Data Science Initiative and Center for Research on Computation and Society, Harvard University [2]University of Colorado, Boulder [3]Microsoft Research [4]Microsoft AI for Good Research Lab [5]School of Computing and Augmented Intelligence, Arizona State University. Correspondence to: Hannah Kerner <hkerner@asu.edu>, Esther Rolf <esther.rolf@colorado.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Machine learning, especially deep learning, was born to extract insights from raw data at scale. However, current ML solutions designed for other data modalities, such as natural images or language, are sub-optimal for satellite data.

Satellite data presents challenges and opportunities distinct from other data modalities (Figure 1). Unlike natural images, the size of targets in satellite images span a logarithmic scale from < 1m (e.g., trees) to > 1km (e.g., forests). Temporal patterns in satellite time series also span logarithmic scales, from hours or days (e.g., floods) to years or decades (e.g., sea level rise). Data are acquired using a variety of sensors that capture diverse spectral channels (beyond 3-channel RGB) and precise measurements (beyond 8 bits). Satellites collect data over the entire surface of the Earth at fixed time intervals and spatial resolutions. Observations are acquired from an overhead perspective from fixed altitudes and lack a "natural" orientation, unlike natural images.

While there has been increasing interest in ML for satellite data (SatML) (Zhu et al. (2017); Table A1), SatML research as a whole falls short on these challenges and opportunities. The mainstream approach to SatML has been to adapt or "lift and shift" solutions designed for other modalities, especially natural images, to satellite data with minimal tailoring. Rather than tackling novel or outstanding challenges, many studies propose ML solutions to well-resolved or low-priority problems in the field of remote sensing (Tuia et al., 2021). Researchers favor large, "state of the art" models designed for other modalities without regard for the unique characteristics and deployment considerations of SatML models. This comes at a cost of performance, scalability, and missed research opportunities.

The unique properties of satellite data and the challenges of SatML—as with other data modalities such as natural images and text—warrant specialized methods for collecting, modelling, and interpreting data. For example, self-supervised learning techniques have been proposed to explicitly learn spatiotemporal relationships in geospatial data, such as nearby locations being more similar than distant locations (Jean et al., 2019) or the same location varying in appearance over different seasons (Manas et al., 2021). Marcos et al. (2018) proposed a small rotation-equivariant neural network that outperformed a larger convolutional

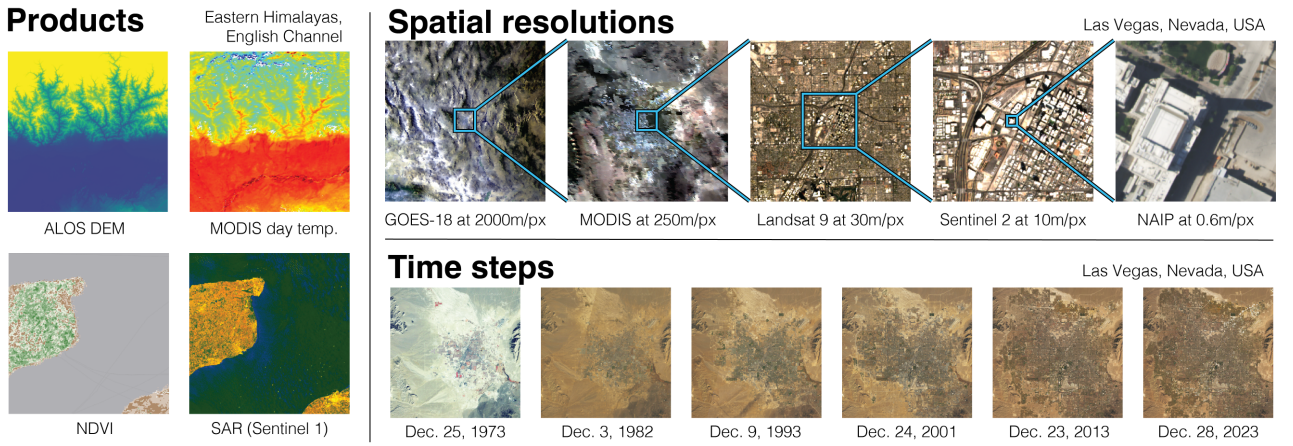

Figure 1. **Satellite images of the same location can vary widely** depending on factors like spatial resolution and cropping extent, temporal dimension, and satellite mission or instrument. ML methods that leverage these factors can drastically outperform methods for general images.

neural network in SatML land cover mapping tasks. While these examples illustrate the potential for specialized SatML approaches, we need to encourage more such research by placing greater value as a community on efforts that tackle the distinct challenges and opportunities of satellite data.

In this position paper, we argue that satellite data constitutes a distinct data modality for ML and that we must recognize it as such to advance the quality and impact of SatML research. Approaches designed for other ML data modalities are ill-fitted for the distinct challenges and opportunities of modeling satellite data and deploying SatML systems (§2). We must prioritize the development of specialized SatML approaches rather than limiting ourselves to "lifting and shifting" existing approaches (§3). Leveraging satellite data can also enrich cross-cutting ML research topics such as distribution shift, self-supervised learning, and multi-modal learning (§4). Realizing the full potential of SatML will require substantial shifts in priorities and practices that the ML community must discuss and act on together (§5).

Prior survey papers detail ML problem settings and applications for remote sensing data (Zhu et al., 2017), promising research opportunities at the intersection of Earth observation and ML (Tuia et al., 2023; 2021; Reichstein et al., 2019; Bergen et al., 2019; Rolnick et al., 2022), and data-centric ML analyses for remote sensing data (Roscher et al., 2023). The goal of this position paper is to present arguments to the general ML community about the opportunities and distinctions of SatML within ML at large. We hope that our arguments provide a path forward – as well as a warning call – for researchers who seek to innovate in SatML or leverage satellite data in cross-cutting ML research.

Many of our arguments for satellite data extend naturally to the broader classes of remote sensing and geospatial data (e.g. aerial images, ERA-5 climate data). We focus on here on SatML, with implications for to the connected sub-fields of ML with remote sensing and geospatial ML.

2. Satellite Data is a Distinct Modality

SatML data have distinct characteristics that are not addressed by ML methods designed for other data modalities.

2.1. Satellite data has unique characteristics

Satellite data are a type of remotely sensed data encompassing diverse datasets with unique characteristics (Figure 1). Satellite data are typically stored in a raster format as a tensor with height, width, and channel dimensions. Temporal information may be included by stacking spatially-aligned rasters along a fourth dimension. While this representation shares similarities with images and videos, several characteristics make modeling satellite data fundamentally different.

Spatial and temporal scales *SatML can model target phenomena that span logarithmic scales in the spatial and temporal dimensions* (Figure 1). Objects and landforms can range in size from less than one meter (e.g., trees or cars) to many kilometers (e.g., forests or wildfires). Temporal patterns can manifest over hours (e.g., earthquakes), weeks (e.g., new construction), seasons (e.g., crop cultivation), years (e.g., glacial retreat), and decades (e.g., sea level rise).

Satellite data is acquired from different instruments and orbits that determine the spatial and temporal resolution of a dataset. Spatial resolution describes the size of the area covered by a single pixel on the Earth's surface. Temporal resolution describes the frequency at which observations of the same area are captured. Some satellites capture data on-demand with very high spatial resolution but less frequent temporal coverage. For example, the WorldView-3 satellite captures images

with 0.3×0.3 m/pixel spatial resolution, but each observation covers a small geographic extent and is infrequent in time. At the other extreme, the MODIS satellite captures 500×500 m/pixel images daily with a large geographic extent.

Spectral channels *SatML can model many spectral channels with greater diversity and precision than standard 8-bit RGB images* (Figure 1). Different satellite sensors capture light from different regions of the electromagnetic spectrum. Passive sensors like optical and thermal imagers measure light reflected or emitted at different wavelengths from the Earth's surface. Active sensors like synthetic aperture radar (SAR), Lidar, and radar altimetry emit radiation and measure the returned signal. For example, Sentinel-1 captures 2-channel (horizontal and vertical polarization) SAR images, while Sentinel-2 captures 13-channel optical images from the visible, near-infrared and shortwave infrared spectrum. Data are typically stored in GeoTIFF or other specialized formats that allow higher radiometric resolution than standard 8-bit image formats like JPEG (e.g., Sentinel-2 has 12-bit resolution).

ML methods developed for 3-channel RGB images are sub-optimal for modeling satellite datasets (§3). In addition, common ML libraries lack support for many-channel satellite images. For example, TorchVision and other libraries with pre-trained models assume images are 3-channel and thus cannot be used for transfer learning with most satellite datasets.

Data volume *SatML methods must efficiently process enormous, continually growing data volumes.* The volume and ever-growing nature of satellite data directly impact the approaches researchers must design to effectively learn from and process the data. Satellite datasets cover the entire globe at high spatial and temporal resolutions. These datasets date back to 1972 and will continue to grow in the future. As of January 2024, the European Space Agency's Copernicus Data Space Ecosystem contained 66 petabytes of public satellite data from the Sentinel missions (Copernicus, 2024). The upcoming NASA-ISRO Synthetic Aperture Radar (NISAR) mission will generate up to 140 petabytes of data (85 TB/day) (Blumenfeld, 2017). These petabyte-scale datasets must be processed efficiently at inference time (§2.2) and present data curation challenges at training time.

Satellite data archives are much larger than existing large datasets used in ML research. For example, the text dataset of snapshots from the Common Crawl database used to train GPT-3 was 45 TB (and 570 GB after filtering) (Brown et al., 2020). LAOIN-5B, the largest paired text-image dataset today, is $\approx$220 TB (Schuhmann et al., 2022). The ILSVRC 2012 ImageNet dataset (Deng et al., 2009), widely used for vision model benchmarking and pre-training, is $\approx$150 GB.

Annotations *SatML target variables are defined with respect to positions (in space and time) on Earth – each of which may be paired with multiple satellite observations. At the same time, label datasets are often sparse with variable schemas and quality.* In contrast to many ML settings where observations, X, and labels, y, are definitive pairs (e.g., X = a photo and y = a label/interpretation of that photo), in SatML, label annotations are often generated irrespective of satellite observations. Instead, labels can be paired with many different choices of satellite observations corresponding to the label's location and time index (Figure 2). For example, satellite imagery can be paired with vector labels from government agricultural GIS databases to create ML ready datasets for different tasks (Schneider et al., 2023; Garnot & Landrieu, 2021; Kondmann et al., 2021). Factors like the spatial resolution, season, and time of day of data acquisition significantly impact the observation and must be considered when modeling the data (Tarasiou & Zafeiriou, 2022).

"Ground-referenced" labels are obtained by a physical visit to a specific location to obtain an annotation indexed by geographic coordinates and time (Bromberg Gaber, 2020). These often involve careful sampling informed by spatio-temporal statistics and are generally expensive and difficult to acquire. As a result, ground-truth datasets are typically small, sparse, spatio-temporally clustered, and specialized in a way that can make harmonization with other datasets difficult (Maskey et al., 2020).

Even when labels are "image-referenced" (e.g., created from expert annotations of satellite observations), severe sampling biases can occur (Reichstein et al., 2019). Several large-scale datasets have increased global coverage of annotated image datasets, such as Functional Map of the World (Christie et al., 2018) and BigEarthNet (Sumbul et al., 2019), but are still more densely sampled in highly populated areas in the Global North. For a thorough review of datasets in the broader machine learning for earth observation space, see (Schmitt et al., 2023) and (Xiong et al., 2022), or repositories such as the IEEE Geoscience and Remote Sensing Society's data platform (Schmitt et al., 2022).

2.2. Deployment challenges of dense prediction

For most deployment settings of SatML, the desired output is a set of dense predictions that form a map of an entire geographic region (Figure 3). For example, crop type classification models generate a predicted map of every pixel in which a specific crop is growing (Nakalembe & Kerner, 2023; Jo et al., 2022), and disaster response maps assess damage across every pixel in an affected area (Xu et al., 2019). [1]

The need for dense prediction in deployment makes model

[1] This is related to the concept of image mosaicing, but predicted labels (not input images) are stitched together over a larger region.

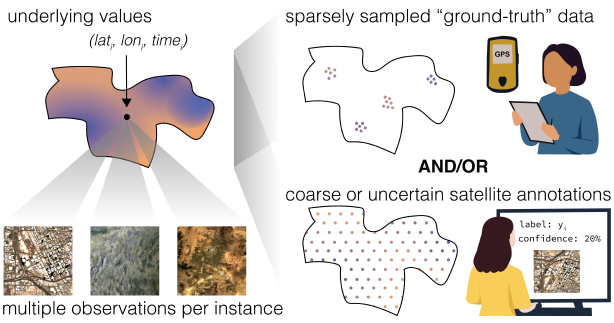

Figure 2. **In SatML, multiple observations and multiple (or no) labels may correspond to a given (lat, lon, time) index**, whereas in many ML settings, labels are defined directly from images.

efficiency an important SatML consideration. For context, at 10m/pixel resolution, the world's land area is represented by nearly 1.5 trillion pixels. Depending on the task requirements, predictions over such data volumes may need to be updated frequently (e.g., sub-weekly for deforestation monitoring (Diniz et al., 2015)). Such prediction volumes can be expensive and time-consuming. Model efficiency is thus one of the foremost deployment considerations for end-users (Van Tricht, 2021; Hengl et al., 2017; Robinson et al., 2019).

Dense prediction also presents opportunities. Models can be applied to overlapping extents or fields-of-view, resulting in multiple predictions per pixel that can be averaged for more consistent maps (Huang et al., 2018). Dense predictions from SatML models can easily be visualized as a larger map, rather than as a set of individual predictions. This can illuminate failure modes for predictive models in a way that is not possible in natural image datasets (Zvonkov et al., 2023).

2.3. Evaluation challenges of spatio-temporal ML

Traditional ML evaluation practices fall short of the needs for SatML. Many recent SatML papers follow a conventional uniform at random assignment of data to training and test sets. While there are spatial interpolation tasks where in-sample performance is relevant, out-of-sample use cases are often the ones where SatML models are most valuable (Rolf, 2023). In these settings, uniformly at random sampled test splits can be a serious limitation to understanding model performance. For example, if labels are spatially auto-correlated and clustered, assigning data instances to training and evaluation splits uniformly at random will likely overestimate model performance in regions outside of available clusters.

Spatially aware holdout and cross-validation methods have been designed to test how models might perform outside the regions in which they were trained. These include blocking or buffering the distance between a train and test set (Figure 3) or parametrically varying the distance between train and test

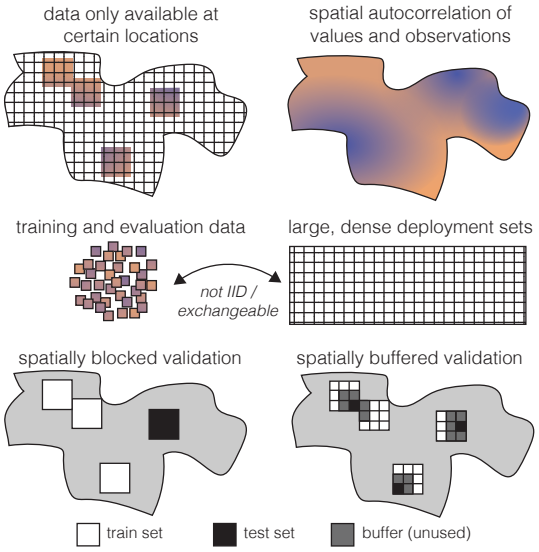

Figure 3. **SatML has distinct considerations for deployment and evaluation.** Deployment datasets are often dense, and much larger than training datasets. Spatio-temporal covariate shifts necessitate spatially aware model validation for out-of-sample model deployment. Figure adapted in part from Zvonkov et al. (2023).

data (Roberts et al., 2017; Le Rest et al., 2014; Pohjankukka et al., 2017; Airola et al., 2019; Wang et al., 2023a).

Estimating the actual accuracy of a predicted map requires carefully sampled ground truth data to measure the predictions against (Wadoux et al., 2021; Stehman & Foody, 2019). Yet, high-quality and uniformly sampled data is scarce (§2.1). The design and sampling of datasets used to evaluate predictions and derived maps over time and space is thus a major endeavor (Olofsson et al., 2014; Tsendbazar et al., 2021).

Assessing whether published SatML models and maps add value (i.e., predictions are accurate, calibrated, and better than random) remains a significant hurdle to developing reliable SatML for real-world settings. SatML inherits many evaluation challenges of geospatial ML, such as (i) reliably estimating accuracy and uncertainty from cluster-sampled reference data (Milà et al., 2022; Wang et al., 2023a) and (ii) developing notions of "regions of applicability" to describe limits on where models are useful (Meyer & Pebesma, 2022). Given the distinct deployment (§2.2) and ethical considerations (§2.4) of SatML models, there is also a need for dedicated performance metrics that measure computational efficiency, privacy, transparency, usability, and fairness.

2.4. Distinct ethical concerns of SatML

While many satellite data products have been in the public domain for decades (§2.1), SatML systems increase the scale and ease at which people and their activities can be monitored. In many countries, property laws protect land owners from

physical invasion of privacy or unsanctioned use of land by others, but these laws do not clearly extend to satellite representations of the same space. *People have no clear way to consent to or opt out of the collection of satellite data or the development of SatML models that cover their land or community*.

SatML is characterized by a "duality" of potential impacts. SatML technologies that can aid efforts in environmental and social justice by providing evidence (Ovienmhada et al., 2023; Sefala et al., 2021) can also be used to "data-fy" unjust practices like redlining, as Sefala et al. (2021) note. Systems that forecast crop yields can improve crop insurance access (Nakalembe & Kerner, 2023), but inaccurate yield predictions can worsen outcomes for farmers (Benami & Carter, 2021). SatML models used for damage assessment after natural disasters also have military applications.

Existing efforts from responsible and ethical ML need significant adaptation to address the distinct concerns of SatML (Kochupillai et al., 2022). For example, a model card template (Mitchell et al., 2019) could be specialized for SatML models to highlight potential for disparate performance across geographies (Aiken et al., 2023), delineate boundaries of intended use (Meyer & Pebesma, 2022), and encourage use-contextualized model evaluations (§2.3).

However, we need to do more than adapt existing efforts. The SatML community must collaborate with stakeholders about the trade-off between the risks and benefits of SatML technologies and understand how decisions made using SatML-generated maps can impact policy (Sandker et al., 2021) and communities (Szantoi et al., 2020). As one step toward this, SatML efforts can integrate CARE principles for indigenous data governance (Carroll et al., 2020) and account for the unique context of satellite data (Dogan & Wood, 2023).

3. SatML Warrants Specialized ML Methods

Since satellite data can often be easily formatted as 3-channel images (e.g., by removing all but 3 spectral channels), many existing computer vision models can be applied "out of the box" to satellite data. While this "lift and shift" strategy can facilitate rapid exploration of research ideas and failure modes, it is also constraining, as existing ML approaches are ill-fitted to many of the distinctions of satellite data (§2).

On the other hand, the SatML characteristics outlined in §2 can drive novel ML algorithms, architectures, and models. In contrast to the mainstream "lift and shift" approach, the following examples evidence how emerging SatML approaches designed explicitly for the satellite data modality have increased performance or real-world applicability over traditional ML approaches. *We argue that the ML community must pursue and catalyze more such research that is designed for the needs of SatML*, rather than being content with methods that work for some tasks despite those needs.

3.1. Learning strategies

ImageNet pre-trained models are popular for transfer learning tasks. However, ImageNet only contains 3-channel (RGB) natural images. Naively adapting ImageNet pre-trained models for satellite data—for example, by using only the RGB channels of a multispectral image—is a clear example of "lifting and shifting" methods designed for natural images to satellite data. While ImageNet models have worked well for some SatML tasks after fine-tuning (Lacoste et al., 2023), they are much worse than random initialization for others (Wang et al., 2022). Bastani et al. (2023) showed that pre-training supervised models with satellite datasets improved performance by as much as 18% over ImageNet pre-training.

Representation learning methods designed specifically for SatML have also outperformed "lift and shift" approaches. Feature extraction using random convolutional features (Rolf et al., 2021; Corley et al., 2023) and random pixel sets (Garnot et al., 2020) are strong baselines for many SatML tasks, but generally are not for natural images. Rußwurm et al. (2024) proposed an adaptive meta-learning model designed for satellite data with a variable number of spectral channels and output classes. Tseng et al. (2022) incorporated task-specific metadata like geographic region in the meta-learning process.

Self-supervised models pre-trained with Landsat or Sentinel-2 multispectral imagery have outperformed ImageNet weights in diverse SatML tasks (Wang et al., 2023b; Stewart et al., 2023). SSL models customized to the unique characteristics of satellite data such as the importance of scale (Reed et al., 2023), temporal dimension (Tseng et al., 2023; Manas et al., 2021), and availability of multiple sensors (Fuller et al., 2023) have outperformed approaches designed for natural images. Cong et al. (2022) proposed a specialized encoding for groups of spectral bands to preserve relevance between individual spectral channels. Ayush et al. (2021) added an auxiliary geolocation prediction task to a similar contrastive pre-training setup. Tseng et al. (2023) and Hackstein et al. (2024) showed that, in a masked autoencoder framework, structured masking strategies outperformed conventional random masking when evaluated on downstream tasks.

3.2. Model architectures

To advance progress on SatML tasks, ML researchers need to re-consider the explicit and inductive biases present in popular deep learning model architectures.

For example, common deep learning architectures designed for natural image benchmarks are translation equivariant, but not rotation equivariant. This design choice stems from the fact that orientation in natural images is often important for preserving object context—predicting the "correct" orientation of rotated images is even an effective pretext task for pre-training (Gidaris et al., 2018). In contrast, objects in

overhead satellite images do not have a "natural" orientation. Considering the special characteristics of satellite imagery, Marcos et al. (2018) showed that rotation equivariant models improve performance in SatML land cover mapping tasks. Van Etten (2018) proposed specialized augmentations in satellite data for rotation invariance, as well as an architecture specialized to detect small, dense clusters of objects common in high-resolution satellite images (e.g., cars in parking lots).

Modern semantic segmentation models such as U-Net (Ronneberger et al., 2015) or DeepLabV3+ (Chen et al., 2018) have large theoretical receptive fields, whereas some SatML tasks, such as land cover mapping, only require small receptive fields (if you assume land cover at a point is a function of the local spectral responses and textures at that point). Brown et al. (2022) leveraged this constraint and other satellite data properties to achieve state-of-the-art land cover segmentation results with an architecture nearly $100\times$ smaller than standard segmentation architectures.

While most computer vision research uses datasets captured by a single sensor type (RGB photographs), satellite data is marked by diversity in sensors. Data fusion research (Steinhausen et al., 2018; Van Tricht et al., 2018; Whyte et al., 2018; Denize et al., 2018; Tseng et al., 2023; Babaeian et al., 2021) has shown that models designed to leverage data from multiple complementary sensors can substantially outperform models trained with data from a single sensor.

3.3. Explicitly modeling domain context

Specialized SatML methods can encode geographic context and constraints. Spatial and spatio-temporal structures are omnipresent in satellite data and can inform model design and training procedures, or may provide inductive biases.

For example, Klemmer et al. (2022) and Klemmer & Neill (2021) incorporated autocorrelation measures into the model loss which improved performance in predictive and generative modeling of geospatial patterns. Jean et al. (2019) used inspiration from Tobler's first law of geography, that "everything is related to everything else, but nearby things are related more", to design an unsupervised triplet loss based on geographic distance between samples. Mac Aodha et al. (2019) and Mai et al. (2023) directly incorporated geographic priors into image classification models by conditioning model outputs on image locations.

Objects in temporal sequences of satellite images typically remain in a fixed position but change in appearance over time (unlike many videos). Tarasiou et al. (2023) designed a method for spatio-temporal encoding in vision transformers. Robinson et al. (2021) leveraged the fact that many human-made structures do not change geographic position once constructed to design a parameter-free change point detection method for satellite time series. Nyborg et al. (2022) proposed embedding temporal dimension by thermal signature to enhance generalizability of agricultural SatML models across regions with different seasonal patterns.

4. Satellite Data Enriches ML Research

By positioning satellite data as a distinct modality alongside other popular modalities like images and text, the entire field stands to make advances in the robustness, scalability, and flexibility of ML methods. The geographic grounding, temporal regularity, global coverage, and sensor diversity of satellite data can inspire the creation of new datasets, experiments, and research directions, and catalyze progress on long-standing research questions. As evidence, we highlight several "core ML" research areas that satellite data can enrich.

4.1. Distribution shift

ML models that are reliable, fair, and safe for real-world use must perform well across a variety of unseen distributions. Previous work in computer vision showed that ML models are highly sensitive to distribution shifts, even those that may be subtle or undetectable by humans (Taori et al., 2020).

Distribution shifts are inevitable and pervasive in SatML contexts (Figure 4). Covariate shifts occur across time, space, spectrum, and scale (§2.1). Concept drift occurs as environmental and social phenomena shift dynamically over time. Previous work suggests that ML models may generalize worse for satellite data than natural images. Benson & Ecker (2020) show that the accuracy of models trained to detect post-disaster building damage in satellite images dropped by as much as 30% when evaluated on a new disaster site. Field delineation models transfer poorly across countries due to differences in farming practices, crop types, and growing seasons (Kerner et al., 2023).

The different wavelengths, temporal cadences, and spatial resolutions of satellite data (§2.1) have driven research in domain adaptation (Tuia et al., 2016; Song et al., 2019). Studying distribution shifts in satellite data provides immense value for SatML use cases and increase the real-world usability of models. It can also enrich ML research by providing new datasets, case studies, and contexts for studying distribution shift and distributional robustness. For instance, the long historical archive of satellite data with global coverage provides a rich data source for studying distribution shift.

4.2. Self-supervised learning and evaluation

Self-supervised learning (SSL) is especially important for SatML, where unlabeled data abounds but labeled data is scarce and difficult to obtain (§2.1). In §3.1, we gave examples of effective SSL strategies designed specifically for satellite data. In this section, we discuss how the unique context of SatML can enrich ML research on self-supervised learning.

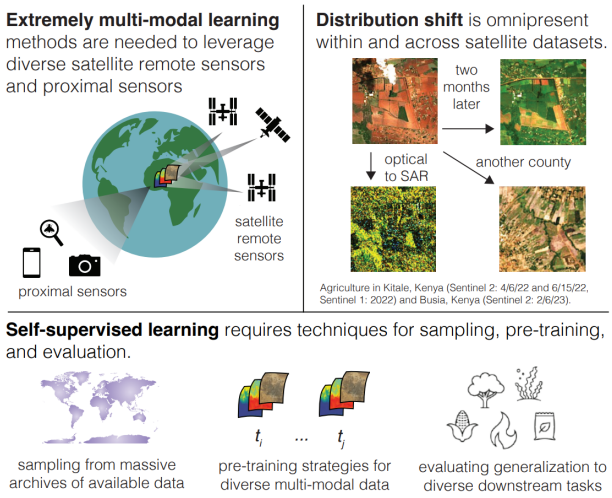

Figure 4. **SatML can enrich many research areas in ML**, e.g., multi-modal, self-supervised, and distributionally robust learning.

The nature of satellite data motivates new SSL research questions (Figure 4). Since the availability, sensor types, and fidelity of satellite observations can vary greatly (§2.1), SatML is an excellent testbed for studying how to extract representations from complex, real-world data. The diverse applications of SatML motivate SSL models capable of generalizing to a wide range of tasks (Lacoste et al., 2023; Yeh et al., 2021; Klemmer et al., 2023). The sheer scale of satellite data raises questions about how to sub-sample data effectively for pre-training, and how data complexity relates to model capacity.

Beyond training, there has been limited research on approaches for *evaluation* without labels. In SatML settings, it is rarely (if ever) feasible to acquire new labeled datasets for every new deployment setting (i.e., for new every year, for target, and region a model is applied in). This practical need adds urgency, real-world relevance, and new challenges to the emerging field of unsupervised, self-supervised, and semi-supervised evaluation (Guillory et al., 2021; Deng & Zheng, 2021; Yu et al., 2022; Baek et al., 2022).

4.3. Multi-modal learning

Piecing together the most complete view of the Earth requires leveraging multiple data sources. Satellite platforms collect and geographically align a variety of observations, including traditional cameras, thermal images, radar or synthetic aperture radar, and Lidar (Figure 1). Derived products like digital elevation models (DEMs) and soil maps (Poggio et al., 2021) can provide a semantic view of ground conditions. Despite this wealth of available sensors, few SatML methods today leverage inputs from more than 1-2 sensors (§3.2) – this is a promising and necessary direction for future ML research.

Satellite data brings rich, real-world challenges for multi-modal ML. For example, fusing high-resolution imagery with low-resolution and partially-observed derived products requires multi-modal learning with coarse or weak observation modalities (Rolf et al., 2022; Li et al., 2022). Datasets such as SEN12MS (Schmitt et al., 2019) or CropHarvest (Tseng et al., 2021) that pair labels with observations from multiple satellites can serve as benchmarks for new algorithms. There are also new, untapped opportunities for multi-modal learning combining raster (image-like) and vector (polygon-like) datasets (Cepeda et al., 2023) that would enrich the scope of multi-modal learning studied by ML researchers today.

Satellite data can complement learning from other modalities including text (Uzkent et al., 2019), natural images (Zhai et al., 2017; Dhakal et al., 2023), sound (Khanal et al., 2023), and other on-the-ground sensors ("proximal sensors") by providing additional supervision or training signal (Tuia et al., 2021; Lefevre et al., 2017). The SatML task of cross-view matching also poses a unique multi-modal learning challenge in which the goal is to match ground-level images with their location in corresponding satellite images (Hu et al., 2018).

4.4. Positional encodings

Positional encodings enable models to learn short- or long-distance interactions in naturally hierarchical data such as text (sequences) or images. They are an integral component in the success achieved by Transformer architectures (Vaswani et al., 2017). Satellite data contain additional spatial and temporal hierarchies and are distributed in a non-Euclidean, spherical space.

Encoding the spatio-temporal position of satellite data in neural network architectures is a distinct research challenge. In recent years, studies on longitude/latitude location encoders for neural networks proposed sinusoidal (Mac Aodha et al., 2019; Mai et al., 2020) and spherical harmonics (Rußwurm et al., 2024) representations. Spatio-temporal location encoders or location encoder pretraining are promising directions for future research. Future work could also investigate encoding additional satellite metadata, such as spatial resolution or sensor angles.

4.5. Human-in-the-loop and active learning

Since labeled data is generally rare but unlabeled data is abundant in SatML, iterating through phases of selective labeling and model retraining can be a promising approach (Ortiz et al., 2022; Desai & Ghose, 2022; Robinson et al., 2020). *The unique nature of SatML raises research questions for human-in-the-loop and active learning* such as how to scale existing active learning query methods to petabyte scale unlabeled datasets. It also brings new constraints: e.g., some geographic areas are unsafe to visit, or more expensive to get to.

5. Synthesis and Discussion Topics

At first glance, SatML may seem like a straightforward application area of ML or computer vision. We argue that this perspective greatly undervalues the distinct challenges and potential of SatML research and that we must recognize satellite data as a distinct modality for ML research. However, much is at stake in intensifying SatML research without community discussion and consensus on how we will orient and measure research progress. Application without understanding—of the data modality, the downstream tasks, or the stakeholders—will not result in real-world impact.

Recognizing and approaching satellite data as a distinct modality in ML will help facilitate focused and responsible exploration in SatML and ML at large. ML researchers aiming to innovate in the sub-field of SatML must address the unique nature of satellite data and SatML problem settings (§2) to design approaches that are purpose-driven, efficient, and impactful (§3). For researchers who work primarily in other areas of ML, SatML data and contexts can amplify research impact and open opportunities for synergistic developments bridging theoretical and applied ML (§4).

The ML community has benefited from developing specialized approaches for distinct data modalities. For example, the convolutional neural network architecture arose from the need to detect objects with translation and illumination invariance in natural images (Krizhevsky et al., 2012). The Transformer architecture arose from the need to capture patterns and dependencies in the language modality (Vaswani et al., 2017). The widely-used U-Net architecture was inspired by the needs of biomedical data (Ronneberger et al., 2015). We have not yet discovered what is possible for satellite data.

In addition to addressing the research challenges and opportunities discussed throughout this paper, realizing the full potential of SatML will require substantial shifts in the status quo for the wider ML community. Towards this, we pose the following discussion points and action items for the community:

How will we prioritize and collaborate on the most important SatML challenges? SatML has the potential to transform many research areas – both within and outside ML – with different norms and institutions. Organizing and supporting researchers from these different communities requires coordinated action.

We must develop an inclusive SatML research community with common priorities and guardrails (Figure 5). SatML must integrate perspectives from many communities and experts within ML and fields like remote sensing, geography, the geosciences, and social sciences. Otherwise, we run the risk of reinventing results that other disciplines have established decades ago, and applying computationally intensive methods for no real value added (Tuia et al., 2021).

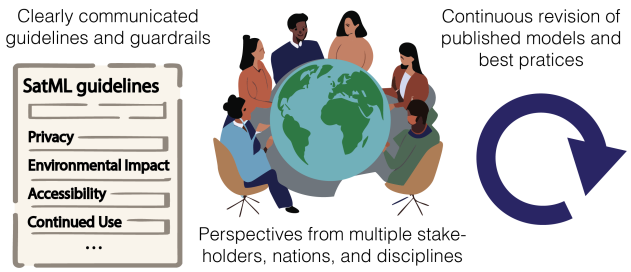

Figure 5. **Developing priorities and guidelines for SatML research** connects many communities, stakeholders, and disciplines.

Disciplinary institutions must acknowledge SatML's place within the scope of ML and computer science research. The key institutions that facilitate research progress (e.g., publishing venues and funding agencies) dictate in large part how SatML research will proceed. With the help of the SatML community, publishing venues must establish what research fits (i) "core ML" venues (e.g., ICML), (ii) venues in other disciplines (e.g., remote sensing journals), and (iii) cross-cutting venues (e.g., remote sensing workshops at core ML venues). Funding agencies must support individual SatML research agendas that advance research beyond the status quo, which in aggregate can influence discipline-level paradigm shifts.

How will we align SatML research progress and real-world impact? Historically, ML progress has been catalyzed by benchmark datasets and challenges: where there is data, there is an opportunity to climb the leaderboard. In contrast, SatML is often motivated by on-the-ground information needs (e.g., from an end-user or research partner who will use the produced model). For SatML, we must reconsider how we abstract core challenges into benchmarks and how we measure research progress and contributions.

SatML benchmarks must be designed to catalyze impact. Massive volumes and varieties of satellite datasets are available, however it is still unclear which of the existing benchmarks and metrics are the best proxy for impact. Ideally, SatML benchmarks and metrics will (i) emphasize the challenges that require advances in core ML (§4) and (ii) realistically capture real-world data and deployment conditions (§2).

We need to emphasize SatML research in the real world. Studies on deploying SatML in on-the-ground conditions will not propose a new algorithm or dataset, but their results can add critical knowledge about the performance or utility (or failure points) of SatML in real-world conditions. We must synthesize findings from deployed SatML research to revisit our established practices, benchmarks, and research directions. We could explicitly value and catalyze such research by designing SatML-specific "impact challenges" or other community milestones to orient research toward

long-term real-world impact (Wagstaff, 2012).

How will we ensure SatML progress benefits global and local communities? When choosing interesting SatML research directions, ML researchers must consider how their research can impact real communities and ecosystems.

We must value global and community-focused research. Some of the biggest potential for SatML impact lies in low-income or under-resourced countries and communities. To benefit from SatML solutions, many end-users need efficient, low-resource ML models due to compute access, internet cost/speed, and other constraints (Nakalembe & Kerner, 2023). The ML community should value research contributions that make SatML technologies accessible to end-users globally by addressing these constraints (Rolf et al., 2021), rather than scoping them as outside of the research domain.

We must re-think our open science practices. The standard practice of openly sharing data, code, and models has been fundamental to rapid progress in ML research. However, we must question this practice when data represent human communities and activities and models can be used for unintended purposes (§2.4). We must create new ways of enabling research progress while minimizing harm and respecting communities, for example through new models of data stewardship (Walter et al., 2021) and data sharing (Sefala et al., 2021).

Impact Statement

The use of machine learning methods for and with satellite data has broad and varying societal implications. As part of our position paper, we argue for including reflections and analyses of distinct social and ethical considerations within the scientific field of SatML (see e.g. §2.4). Like other application-focused subdisciplines of machine learning, SatML is characterized by a "duality" of potential impacts: For example, one the one hand, SatML may be used to support environmental protection efforts, while on the other hand it may encroach on privacy of people and communities. As such, the SatML community must integrate discussions on social and ethical implications into its scientific agenda.

Acknowledgements

We thank Marc Rußwurm, Claire Monteleoni, Sherrie Wang, David Rolnick, Anthony Ortiz, Isaac Corley, Patrick Gray, Gabriel Tseng, Brice Ménard, and Max Simchowitz for their valuable feedback on the arguments supporting our position. We thank Mirali Purohit, Rahul Nair, Snehal Chaudhari, Aditya Mohan, and Ivan Zvonkov for feedback on writing and clarity.

Funding. ER is supported by the Harvard Data Science Initiative, the Center for Research on Computation and Society at Harvard, and Microsoft.

References

Aiken, E., Rolf, E., and Blumenstock, J. Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy. *IJCAI*, 2023.

Airola, A., Pohjankukka, J., Torppa, J., Middleton, M., Nykänen, V., Heikkonen, J., and Pahikkala, T. The spatial leave-pair-out cross-validation method for reliable AUC estimation of spatial classifiers. *Data Mining and Knowledge Discovery*, 33(3):730–747, 2019.

Ayush, K., Uzkent, B., Meng, C., Tanmay, K., Burke, M., Lobell, D., and Ermon, S. Geography-aware self-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10181–10190, 2021.

Babaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V. K., and Tuller, M. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. *Remote Sensing of Environment*, 260:112434, 2021.

Baek, C., Jiang, Y., Raghunathan, A., and Kolter, J. Z. Agreement-on-the-line: Predicting the performance of neural networks under distribution shift. *Advances in Neural Information Processing Systems*, 35:19274–19289, 2022.

Bastani, F., Wolters, P., Gupta, R., Ferdinando, J., and Kembhavi, A. SatlasPretrain: A large-scale dataset for remote sensing image understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16772–16782, 2023.

Benami, E. and Carter, M. R. Can digital technologies reshape rural microfinance? Implications for savings, credit, & insurance. *Applied Economic Perspectives and Policy*, 43(4):1196–1220, 2021.

Benson, V. and Ecker, A. Assessing out-of-domain generalization for robust building damage detection. *NeurIPS 2020 Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response*, 2020.

Bergen, K. J., Johnson, P. A., de Hoop, M. V., and Beroza, G. C. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433):eaau0323, 2019.

Blumenfeld, J. Getting ready for nisar—and for managing big data using the commercial cloud. *EarthData: Open Access for Data Science*, 2017. URL https://www.earthdata.nasa.gov/learn/articles/getting-ready-for-nisar.

Bromberg Gaber, Y. A guide for collecting and sharing ground reference data for machine learning applications. *Medium*, 2020. URL https://medium.com/radiant-earth-insights/a-guide-for-collecting-and-sharing-ground-reference-data-for-machine-learning-applications-90664930925e.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., et al. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1):251, 2022.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33: 1877–1901, 2020.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., et al. The care principles for indigenous data governance. *Data Science Journal*, 19:43–43, 2020.

Cepeda, V. V., Nayak, G. K., and Shah, M. GeoCLIP: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Advances in Neural Information Processing Systems*, 2023.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.

Christie, G., Fendley, N., Wilson, J., and Mukherjee, R. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6172–6180, 2018.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D., and Ermon, S. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022.

Copernicus. Copernicus Data Space Dashboard. https://dashboard.dataspace.copernicus.eu/#/service-insight/data-published, 2024. Accessed: January 17th, 2024.

Corley, I., Robinson, C., Dodhia, R., Ferres, J. M. L., and Najafirad, P. Revisiting pre-trained remote sensing model benchmarks: Resizing and normalization matters. *arXiv preprint arXiv:2305.13456*, 2023.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Deng, W. and Zheng, L. Are labels always necessary for classifier accuracy evaluation? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15069–15078, 2021.

Denize, J., Hubert-Moy, L., Betbeder, J., Corgne, S., Baudry, J., and Pottier, E. Evaluation of using Sentinel-1 and-2 time-series to identify winter land use in agricultural landscapes. *Remote Sensing*, 2018.

Desai, S. and Ghose, D. Active learning for improved semi-supervised semantic segmentation in satellite images. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 553–563, 2022.

Dhakal, A., Ahmad, A., Khanal, S., Sastry, S., and Jacobs, N. Sat2Cap: Mapping fine-grained textual descriptions from satellite images. *arXiv preprint arXiv:2307.15904*, 2023.

Diniz, C. G., de Almeida Souza, A. A., Santos, D. C., Dias, M. C., Da Luz, N. C., De Moraes, D. R. V., Maia, J. S., Gomes, A. R., da Silva Narvaes, I., Valeriano, D. M., et al. DETER-B: The new Amazon near real-time deforestation detection system. *IEEE Journal of selected topics in applied earth observations and remote sensing*, 8(7): 3619–3628, 2015.

Dogan, A. L. and Wood, D. "Do you collect data to give to the university or do you do the work to benefit people?": Indigenous data sovereignty in environmental contexts. In *Proceedings of the 6th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, pp. 107–116, 2023.

Fuller, A., Millard, K., and Green, J. R. CROMA: Remote sensing representations with contrastive radar-optical masked autoencoders. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=ezqI5WgGvY.

Garnot, V. S. F. and Landrieu, L. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4872–4881, 2021.

Garnot, V. S. F., Landrieu, L., Giordano, S., and Chehata, N. Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12325–12334, 2020.

Gidaris, S., Singh, P., and Komodakis, N. Unsupervised representation learning by predicting image rotations. In *International Conference on Learning Representations*, 2018.

Guillory, D., Shankar, V., Ebrahimi, S., Darrell, T., and Schmidt, L. Predicting with confidence on unseen distributions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1134–1144, 2021.

Hackstein, J., Sumbul, G., Clasen, K. N., and Demir, B. Exploring masked autoencoders for sensor-agnostic image retrieval in remote sensing, 2024.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, 2017.

Hu, S., Feng, M., Nguyen, R. M., and Lee, G. H. CVM-Net: Cross-view matching network for image-based ground-to-aerial geo-localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7258–7267, 2018.

Huang, B., Reichman, D., Collins, L. M., Bradbury, K., and Malof, J. M. Tiling and stitching segmentation output for remote sensing: Basic challenges and recommendations. *arXiv preprint arXiv:1805.12219*, 2018.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, D., and Ermon, S. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3967–3974, 2019.

Jo, H.-W., Koukos, A., Sitokonstantinou, V., Lee, W.-K., and Kontoes, C. Towards global crop maps with transfer learning. *arXiv preprint arXiv:2211.04755*, 2022.

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., and Chuvieco, E. Towards delivering on the sustainable development goals using Earth observations. *Remote Sensing of Environment*, 247:111930, 2020.

Kerner, H., Sundar, S., and Satish, M. Multi-region transfer learning for segmentation of crop field boundaries in satellite images with limited labels. In *AAAI Conference on Artificial Intelligence (AAAI) Workshops*, 2023. URL https://ai-2-ase.github.io/papers/14%5CSubmission%5CField_boundary_delineation___AAAI_2023_AI2SE-camera-ready.pdf.

Khanal, S., Sastry, S., Dhakal, A., and Jacobs, N. Learning tri-modal embeddings for zero-shot soundscape mapping. *British Machine Vision Conference*, 2023.

Klemmer, K. and Neill, D. B. Auxiliary-task learning for geographic data with autoregressive embeddings. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, pp. 141–144, 2021. URL https://doi.org/10.1145/3474717.3483922.

Klemmer, K., Xu, T., Acciaio, B., and Neill, D. B. SPATE-GAN: Improved generative modeling of dynamic spatio-temporal patterns with an autoregressive embedding loss. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4):4523–4531, Jun. 2022. doi: 10.1609/aaai.v36i4.20375. URL https://ojs.aaai.org/index.php/AAAI/article/view/20375.

Klemmer, K., Rolf, E., Robinson, C., Mackey, L., and Rußwurm, M. SatCLIP: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023.

Kochupillai, M., Kahl, M., Schmitt, M., Taubenböck, H., and Zhu, X. X. Earth observation and artificial intelligence: Understanding emerging ethical issues and opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(4): 90–124, 2022. doi: 10.1109/MGRS.2022.3208357.

Kondmann, L., Toker, A., Rußwurm, M., Camero, A., Peressuti, D., Milcinski, G., Mathieu, P.-P., Longépé, N., Davis, T., Marchisio, G., et al. Denethor: The dynamicearthnet dataset for harmonized, inter-operable, analysis-ready, daily crop monitoring from space. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012.

Lacoste, A., Lehmann, N., Rodriguez, P., Sherwin, E. D., Kerner, H., Lütjens, B., Irvin, J. A., Dao, D., Alemohammad, H., Drouin, A., et al. Geo-bench: Toward foundation models for earth monitoring. *37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks*, 2023.

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., and Bretagnolle, V. Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global ecology and biogeography*, 23(7):811–820, 2014.

Lefevre, S., Tuia, D., Wegner, J. D., Produit, T., and Nassar, A. S. Toward seamless multiview scene analysis from

satellite to street level. *Proceedings of the IEEE*, 105(10): 1884–1899, 2017.

Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., and Chanussot, J. Deep learning in multimodal remote sensing data fusion: A comprehensive review. *International Journal of Applied Earth Observation and Geoinformation*, 112:102926, 2022. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2022.102926. URL https://www.sciencedirect.com/science/article/pii/S1569843222001248.

Mac Aodha, O., Cole, E., and Perona, P. Presence-only geographical priors for fine-grained image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9596–9606, 2019.

Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., and Lao, N. Multi-scale representation learning for spatial feature distributions using grid cells. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rJljdh4KDH.

Mai, G., Lao, N., He, Y., Song, J., and Ermon, S. CSP: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In *Proceedings of the 40th International Conference on Machine Learning*, 2023.

Manas, O., Lacoste, A., Giró-i Nieto, X., Vazquez, D., and Rodriguez, P. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9414–9423, 2021.

Marcos, D., Volpi, M., Kellenberger, B., and Tuia, D. Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models. *ISPRS journal of photogrammetry and remote sensing*, 145:96–107, 2018.

Maskey, M., Alemohammad, H., Murphy, K., and Ramachandran, R. Advancing AI for Earth science: A data systems perspective. *Eos*, 101, 2020.

Meyer, H. and Pebesma, E. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, 13(1):1–4, 2022.

Milà, C., Mateu, J., Pebesma, E., and Meyer, H. Nearest neighbour distance matching leave-one-out cross-validation for map validation. *Methods in Ecology and Evolution*, 13(6):1304–1316, 2022.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., and Gebru, T. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pp. 220–229, 2019.

Nakalembe, C. Urgent and critical need for sub-Saharan African countries to invest in Earth observation-based agricultural early warning and monitoring systems. *Environmental Research Letters*, 15(12):121002, 2020.

Nakalembe, C. and Kerner, H. Considerations for AI-EO for agriculture in Sub-Saharan Africa. *Environmental Research Letters*, 18(4):041002, 2023.

Nyborg, J., Pelletier, C., and Assent, I. Generalized classification of satellite image time series with thermal positional encoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1392–1402, 2022.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57, 2014.

Ortiz, A., Negandhi, D., Mysorekar, S. R., Nagaraju, S. K., Kiesecker, J., Robinson, C., Bhatia, P., Khurana, A., Wang, J., Oviedo, F., et al. An artificial intelligence dataset for solar energy locations in India. *Scientific Data*, 9(1):497, 2022.

Ovienmhada, U., Diongue, A., Pellow, D. N., and Wood, D. Satellite remote sensing for environmental data justice: Perspectives from anti-prison community organizers on the uses of geospatial data. *Environmental Justice*, 2023.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E., and Rossiter, D. Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil*, 7(1):217–240, 2021.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., and Heikkonen, J. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, 31(10): 2001–2019, 2017.

Reed, C. J., Gupta, R., Li, S., Brockman, S., Funk, C., Clipp, B., Keutzer, K., Candido, S., Uyttendaele, M., and Darrell, T. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4088–4099, October 2023.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, f. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204, 2019.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

Robinson, C., Hou, L., Malkin, K., Soobitsky, R., Czawlytko, J., Dilkina, B., and Jojic, N. Large scale high-resolution land cover mapping with multi-resolution data. In *CVPR*, 2019.

Robinson, C., Ortiz, A., Malkin, K., Elias, B., Peng, A., Morris, D., Dilkina, B., and Jojic, N. Human-machine collaboration for fast land cover mapping. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 2509–2517, 2020.

Robinson, C., Ortiz, A., Ferres, J. M. L., Anderson, B., and Ho, D. E. Temporal cluster matching for change detection of structures from satellite imagery. In *ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 138–146, 2021.

Rolf, E. Evaluation challenges for geospatial ML. *ICLR Workshop on Machine Learning for Remote Sensing*, 2023.

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., and Hsiang, S. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1):4392, 2021.

Rolf, E., Malkin, N., Graikos, A., Jojic, A., Robinson, C., and Jojic, N. Resolving label uncertainty with implicit posterior models. In Cussens, J. and Zhang, K. (eds.), *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 1707–1717. PMLR, 01–05 Aug 2022.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55(2):1–96, 2022.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, 2015.

Roscher, R., Rußwurm, M., Gevaert, C., Kampffmeyer, M., Santos, J. A. d., Vakalopoulou, M., Hänsch, R., Hansen, S., Nogueira, K., Prexl, J., et al. Data-centric machine learning for geospatial remote sensing data. *arXiv preprint arXiv:2312.05327*, 2023.

Rußwurm, M., Wang, S., Kellenberger, B., Roscher, R., and Tuia, D. Meta-learning to address diverse Earth observation problems across resolutions. *Communications Earth & Environment*, 5(1):37, 2024.

Rußwurm, M., Klemmer, K., Rolf, E., Zbinden, R., and Tuia, D. Geographic location encoding with spherical harmonics and sinusoidal representation networks. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=PudduufFLa&noteId=pF29fBqznM.

Sandker, M., Carrillo, O., Leng, C., Lee, D., d'Annunzio, R., and Fox, J. The importance of high–quality data for REDD+ monitoring and reporting. *Forests*, 12(1):99, 2021.

Schmitt, M., Hughes, L. H., Qiu, C., and Zhu, X. X. Sen12ms – a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W7:153–160, 2019.

Schmitt, M., Ghamisi, P., Yokoya, N., and Hänsch, R. EOD: The IEEE GRSS Earth Observation Database. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5365–5368. IEEE, 2022.

Schmitt, M., Ahmadi, S. A., Xu, Y., Taşkın, G., Verma, U., Sica, F., and Hänsch, R. There are no data like more data: Datasets for deep learning in earth observation. *IEEE Geoscience and Remote Sensing Magazine*, 2023.

Schneider, M., Schelte, T., Schmitz, F., and Körner, M. Eurocrops: The largest harmonized open crop dataset across the european union. *Scientific Data*, 10(1):612, 2023.

Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35: 25278–25294, 2022.

Sefala, R., Gebru, T., Mfupe, L., Moorosi, N., and Klein, R. Constructing a visual dataset to study the effects of spatial apartheid in South Africa. In *Thirty-fifth conference on neural information processing systems datasets and benchmarks track (round 2)*, 2021.

Skidmore, A. K., Coops, N. C., Neinavaz, E., Ali, A., Schaepman, M. E., Paganini, M., Kissling, W. D., Vihervaara, P., Darvishzadeh, R., Feilhauer, H., et al. Priority list of biodiversity metrics to observe from space. *Nature Ecology & Evolution*, 5(7):896–906, 2021.

Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., and Wang, S. Domain adaptation for convolutional neural networks-based remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 16(8): 1324–1328, 2019.

Stehman, S. V. and Foody, G. M. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, 231:111199, 2019.

Steinhausen, M. J., Wagner, P. D., Narasimhan, B., and Waske, B. Combining sentinel-1 and sentinel-2 data for improved land use and land cover mapping of monsoon regions. *International journal of applied Earth observation and geoinformation*, 2018.

Stewart, A. J., Lehmann, N., Corley, I. A., Wang, Y., Chang, Y.-C., Braham, N. A. A., Sehgal, S., Robinson, C., and Banerjee, A. Ssl4eo-l: Datasets and foundation models for landsat imagery. *arXiv preprint arXiv:2306.09424*, 2023.

Sumbul, G., Charfuelan, M., Demir, B., and Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5901–5904. IEEE, 2019.

Szantoi, Z., Geller, G. N., Tsendbazar, N.-E., See, L., Griffiths, P., Fritz, S., Gong, P., Herold, M., Mora, B., and Obregón, A. Addressing the need for improved land cover map products for policy support. *Environmental Science & Policy*, 112:28–35, 2020.

Taori, R., Dave, A., Shankar, V., Carlini, N., Recht, B., and Schmidt, L. Measuring robustness to natural distribution shifts in image classification. In *Advances in Neural Information Processing Systems*, volume 33, pp. 18583–18599, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/d8330f857a17c53d217014ee776bfd50-Paper.pdf.

Tarasiou, M. and Zafeiriou, S. Deepsatdata: Building large scale datasets of satellite images for training machine learning models. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4070–4073. IEEE, 2022.

Tarasiou, M., Chavez, E., and Zafeiriou, S. Vits for sits: Vision transformers for satellite image time series. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10418–10428, June 2023.

Tsendbazar, N., Herold, M., Li, L., Tarko, A., De Bruin, S., Masiliunas, D., Lesiv, M., Fritz, S., Buchhorn, M., Smets, B., et al. Towards operational validation of annual global land cover maps. *Remote Sensing of Environment*, 266: 112686, 2021.

Tseng, G., Zvonkov, I., Nakalembe, C. L., and Kerner, H. Cropharvest: A global dataset for crop-type classification. In *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=JtjzUXPEaCu.

Tseng, G., Kerner, H., and Rolnick, D. TIML: Task-informed meta-learning for crop type mapping. In *AI for Agriculture and Food Systems*, 2022.

Tseng, G., Cartuyvels, R., Zvonkov, I., Purohit, M., Rolnick, D., and Kerner, H. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023.

Tuia, D., Persello, C., and Bruzzone, L. Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):41–57, 2016.

Tuia, D., Roscher, R., Wegner, J. D., Jacobs, N., Zhu, X., and Camps-Valls, G. Toward a collective agenda on AI for Earth science data analysis. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):88–104, 2021.

Tuia, D., Schindler, K., Demir, B., Camps-Valls, G., Zhu, X. X., Kochupillai, M., Džeroski, S., van Rijn, J. N., Hoos, H. H., Del Frate, F., et al. Artificial intelligence to advance Earth observation: a perspective. *arXiv preprint arXiv:2305.08413*, 2023.

Uzkent, B., Sheehan, E., Meng, C., Tang, Z., Burke, M., Lobell, D., and Ermon, S. Learning to interpret satellite images in global scale using Wikipedia. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019.

Van Etten, A. You only look twice: Rapid multi-scale object detection in satellite imagery. *arXiv preprint arXiv:1805.09512*, 2018.

Van Tricht, K. Mapping crops at global scale! What works and what doesn't? https://blog.vito.be/remotesensing/worldcereal-benchmarking, 2021. Accessed: 2023-07-31.

Van Tricht, K., Gobin, A., Gilliams, S., and Piccard, I. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sensing*, 2018.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Wadoux, A. M.-C., Heuvelink, G. B., De Bruin, S., and Brus, D. J. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, 457: 109692, 2021.

Wagstaff, K. L. Machine learning that matters. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1851–1856, 2012.

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., and Lee, V. Indigenous data sovereignty in the era of big data and open data. *Australian Journal of Social Issues*, 56(2):143–156, 2021.

Wang, D., Zhang, J., Du, B., Xia, G.-S., and Tao, D. An empirical study of remote sensing pretraining. *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

Wang, J., Hallman, T., Hopkins, L., Kilbride, J. B., Robinson, W. D., and Hutchinson, R. Model evaluation for geospatial problems. In *NeurIPS 2023 Workshop on Computational Sustainability: Promises and Pitfalls from Theory to Deployment*, 2023a.

Wang, Y., Braham, N. A. A., Xiong, Z., Liu, C., Albrecht, C. M., and Zhu, X. X. SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3): 98–106, 2023b. doi: 10.1109/MGRS.2023.3281651.

Whyte, A., Ferentinos, K. P., and Petropoulos, G. P. A new synergistic approach for monitoring wetlands using Sentinels-1 and 2 data with object-based machine learning algorithms. *Environmental Modelling & Software*, 2018.

Xiong, Z., Zhang, F., Wang, Y., Shi, Y., and Zhu, X. X. Earthnets: Empowering AI in earth observation. *arXiv preprint arXiv:2210.04936*, 2022.

Xu, J. Z., Lu, W., Li, Z., Khaitan, P., and Zaytseva, V. Building damage detection in satellite imagery using convolutional neural networks. *arXiv preprint arXiv:1910.06444*, 2019.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875–883, 2013.

Yeh, C., Meng, C., Wang, S., Driscoll, A., Rozi, E., Liu, P., Lee, J., Burke, M., Lobell, D. B., and Ermon, S. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Yu, Y., Yang, Z., Wei, A., Ma, Y., and Steinhardt, J. Predicting out-of-distribution error with the projection norm. In *International Conference on Machine Learning*, pp. 25721–25746. PMLR, 2022.

Zhai, M., Bessinger, Z., Workman, S., and Jacobs, N. Predicting ground-level scene layout from aerial imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 867–875, 2017.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine*, 5(4):8–36, 2017.

Zvonkov, I., Tseng, G., Nakalembe, C., and Kerner, H. Openmapflow: a library for rapid map creation with machine learning and remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 14655–14663, 2023.

A. Related workshops

We are aware of the following instances of workshops that have been hosted at CS/ML conferences and that promote SatML research (Table 1).

Workshop Name	Conference Name	Years active
DeepGlobe	CVPR	2018
EarthVision	CVPR	2015, 2019-2024
Object Tracking and Classification in and Beyond the Visible Spectrum	CVPR	2004-2011
Perception Beyond the Visible Spectrum	CVPR	2012-2023
Machine Vision for Earth Observation and Environment Monitoring	BMVC	2023
Complex Data Challenges in Earth Observation	CIKM, IJCAI	2021, 2022
GeoAI	SIGSPATIAL	2019, 2021, 2023
Machine Learning for Remote Sensing	ICLR	2023-2024
Workshop on Machine Learning for Earth Observation (MACLEAN)	ECML/PKDD	2019-2023

Table 1. Workshops among CS/ML conferences that promote SatML related research.