

BURAK EKIM , TIMO T. STOMBERG ,
RIBANA ROSCHER , AND MICHAEL SCHMITT 

MapInWild

A remote sensing dataset to address the question of what makes nature wild

The advancement in deep learning (DL) techniques has led to a notable increase in the number and size of annotated datasets in a variety of domains, with remote sensing (RS) being no exception [1]. Also, an increase in Earth observation (EO) missions and the easy access to globally available and free geodata have opened up new research opportunities. Although numerous RS datasets have been published in the past years [2], [3], [4], [5], [6], most of them addressed tasks concerning man-made environments, such as building footprint extraction and road network classification, leaving the environmental and ecology-related subareas of RS underrepresented. Nevertheless, environmental protection has always been an important topic in the RS community, with RS being a useful tool to support conservation policies and strategies combating challenges such as deforestation and loss of biodiversity [7], [8], [9]. Thus, in this article, to meet the pressing need to better understand the nature we are living in, we introduce a novel task of wilderness mapping and advertise the MapInWild dataset [10]—a multimodal large-scale benchmark dataset designed for the task of wilderness mapping from space.

INTRODUCTION

Automated classification of image data has a long tradition in EO. In this community, the classification task can be addressed in different ways, most notably as a scene- or patch-wise classification in contrast to pixel-wise classification. In scene classification, full scenes are assigned by the classifier with single or multiple class labels, whereas in pixel-wise classification (usually called semantic segmentation by the computer vision community), the task outputs densely annotated prediction maps on a pixel scale by separating the input into distinct and semantically coherent segments.

In general, image classification approaches either are based on hand-crafted feature engineering and subsequent machine learning (ML) models or feature learning incorporated into the ML model in the form of deep neural networks. In the RS area, the long-standing tasks of scene classification and semantic segmentation have been approached in a variety of settings [11], ranging from metric learning [12] to multitask learning [13]. While some methods frame the RS-related tasks within the context of perturbation-seeking generative adversarial networks [14], some others have made use of uncertainty estimation applied to deep ensembles [15] and self-attention context networks under adversarial attacks [16].

The success of DL models comes at the expense of decreased interpretability, which means the ability to understand the decision process of the model and the reason why a specific outcome was derived. This is mainly caused by their formation of hundreds of successive layers, leading to a high number of parameters. In recent years, several studies addressed the lack of explainability and interpretability of DL models and proposed methods to overcome this challenge. Model-agnostic approaches are independent of the used model and can be applied post hoc. Popular approaches are, for example, occlusion sensitivity maps [17], which observe the change in the output while systematically occluding small parts of the input; local interpretable model-agnostic explanations [19], which approximate the model with smaller models with fewer parameters; and gradient-weighted class activation maps [18], which combine the activation maps in a convolutional neural network with class-specific gradients. Besides model-agnostic approaches, model-specific approaches are tailored to specific models. They mainly use parts of the model, such as the weights, to analyze the decision process or to interpret the model outcome. One example is activation space occlusion sensitivity (ASOS), which

Digital Object Identifier 10.1109/MGRS.2022.3226525
Date of current version: 31 March 2023

analyzes the activation maps and utilizes the concept of occlusion sensitivity in the activation space.

The task of wilderness mapping is interesting in a twofold way: First of all, wilderness areas are an essential element of the natural environment and provide native habitat for many species, which are often endangered. Thus, the very task itself is important in the context of conservation and environmental protection. Second, though, the term *wilderness* is of a philosophical nature and is comparably ill defined from a mathematical/technical standpoint [20], [21].

Explainable ML methods can be used to address the task of wilderness mapping. The objective of these methods is to uncover patterns in the decision-making process of the mapping task with the goal of improving the understanding of what makes nature wild. Explainable ML approaches, which are used for the discovery of new scientific

knowledge, such as the exploration of wildlife characteristics, combine interpretation tools that present complex processes as in neural networks in a human-understandable space with domain knowledge to derive explanations [22].

Taking the ML perspective further, we claim that training models for wilderness mapping are not only an ideal testbed for methodical developments in explainable ML but also in other subfields, such as weakly supervised learning. Thus, the dataset will not only cater to the environmental RS community but also address technical audiences beyond that.

WILDERNESS MAPPING

Anthropogenic pressure (i.e., human influence) on the environment is the largest single cause of loss of biological diversity [23], and the conservation of ecological dynamics and biodiversity is now more important than ever before. With the ever-increasing prominence of the need for environmental conservation, the world is in desperate need of environmental monitoring and quantifying of the degree of human influence on the environment.

While there is a lot of work being done in fields such as unsupervised, semisupervised, and weakly supervised ML, most of the established methods still rely on the supervised learning paradigm, rendering the need for carefully annotated training data. In the context of environmental RS, this can be achieved for rather well-defined target classes, such as “water” or “forests,” but it becomes a significant challenge if a less technical interpretation of the environment is the desired goal. As an example, many researchers in the past have presented rule-based approaches to study concepts such as naturalness [24], [25], human influence [26], or wilderness [27]. All of these approaches are carefully designed by domain experts and cannot be validated against

actual, biophysically measurable ground truth. The term “wilderness” is philosophical rather than technical in nature and subject to many different definitions [20], with the definition of the European Commission being [29], p. 10]:

A wilderness is an area governed by natural processes. It is composed of native habitats and species, and large enough for the effective ecological functioning of natural processes. It is unmodified or only slightly modified and without intrusive or extractive human activity, settlements, infrastructure or visual disturbance.

This definition assigns four central ecological aspects to the concept of wilderness: 1) naturalness, 2) undisturbedness, 3) undevelopedness, and 4) scale. While scale can easily be defined in a mathematical sense, e.g., by defining so-called *minimum mapping units*, naturalness refers to ecosystems functioning in a natural way and can thus hardly be measured in a technical sense, but is subject to judgment by ecology experts based on in situ observations. Undisturbedness and undevelopedness, however, are concepts that can potentially be observed with the use of RS technologies. As per the European Commission, undevelopedness refers to the absence of “habitation, settlements or other human artefacts such as power lines, roads, railways, fences [that] may hinder ecological processes directly or by promoting the likelihood of human interference” [29], p. 12).

While the task of wilderness mapping could be considered as a special case of land cover mapping, there is a subtle difference to be noted that appear from the ill-posed definition of the term *wilderness*. Since there are nomenclatures with precise class definitions to follow for conventional land cover mapping, it is relatively easier to decide what class to assign to the instance of interest, especially in the presence of domain knowledge. However, given the vague definition of the term, we argue that wilderness is not an instance to map, but a concept to discover.

It is undeniable that wilderness areas pose an utmost significance given their role in harboring species and habitats, which are indisputable elements of the environment. This feature of the term wilderness gives a rise to incorporating explainable ML methods with the purpose of interpreting the decisions made by the DL.

Wilderness has no biophysical basis and thus cannot be easily delineated or categorized. Instead, wilderness is a cultural concept [30], and the associated definition of wilderness changes over time. Moreover, changing policy and management strategies make the identification and protection of wilderness compelling. There exists a diverse set of explanations of the word “wilderness,” most of which are engaged with describing it in a philosophical way, which might seem vague from a technical point of view. Besides, understanding wilderness—a key element of nature—poses great importance since the protection of nature can be fostered by better defining its ill-defined elements. Consequently, we adopt a bottom-up approach and address the ambiguity in nature’s vital elements (e.g.,

GIVEN THE VAGUE DEFINITION OF THE TERM, WE ARGUE THAT WILDERNESS IS NOT AN INSTANCE TO MAP, BUT A CONCEPT TO DISCOVER.

wilderness) to better protect the environment we live in. Ultimately, we raise the rhetorical question “What makes nature wild?” and invite the community to participate in finding possible solutions to the open challenges we have mentioned to get closer to discovering the concept of wilderness.

THE MAPINWILD DATASET

CURATION

The cloud storage and computing facilities of Google Earth Engine (GEE) [32] are utilized during the dataset curation phase by keeping the reproducibility aspect in mind (i.e., defining fixed seed values to random number generators and using definitive time frames for image queries). Given the complexity of the task being addressed, MapInWild consists of distinct sets of geosensors that aim to leverage complementary information about the natural environment: Sentinel-1, Sentinel-2, European Space Agency (ESA) WorldCover, the Visible Infrared Imaging Radiometer Suite (VIIRS), and World Database on Protected Areas (WDPA) polygons; a sample from MapInWild is shown in Figure 1. WDPA polygons comprise feature classes designating protected areas and formed by the United Nations Environment Programme’s World Conservation Monitoring Center and the International Union for Conservation of Nature

(IUCN) World Commission on Protected Areas [28]. The complete list of WDPA classes is given in Table 1. From the anthropogenic pressure point of view, in which the variations in the diversity are explained along human influence gradients, we reformulate these discrete sets of areas on a continuous spectrum, where the distinction among areas is relatively ill defined. On the grounds of the observation that human presence increases going from category Ia to category II, we raise the question: At what point of the human presence continuum should an area be considered a wilderness area? Supported by this theoretical research question, we argue that category Ia and category II could be beneficial proxies in interpreting and explaining the concept of wilderness. Ultimately, we form the MapInWild dataset with the following three WDPA classes: strict nature reserve (category Ia), wilderness area (category Ib), and national park (category II).

As for the Sentinel-1 image, a ground-range-detected product with a resolution of 10 m is used. Sentinel-1 images contain VH and VV polarizations and use the interferometric

WE ADOPT A BOTTOM-UP APPROACH AND ADDRESS THE AMBIGUITY IN NATURE’S VITAL ELEMENTS (E.G., WILDERNESS) TO BETTER PROTECT THE ENVIRONMENT WE LIVE IN.

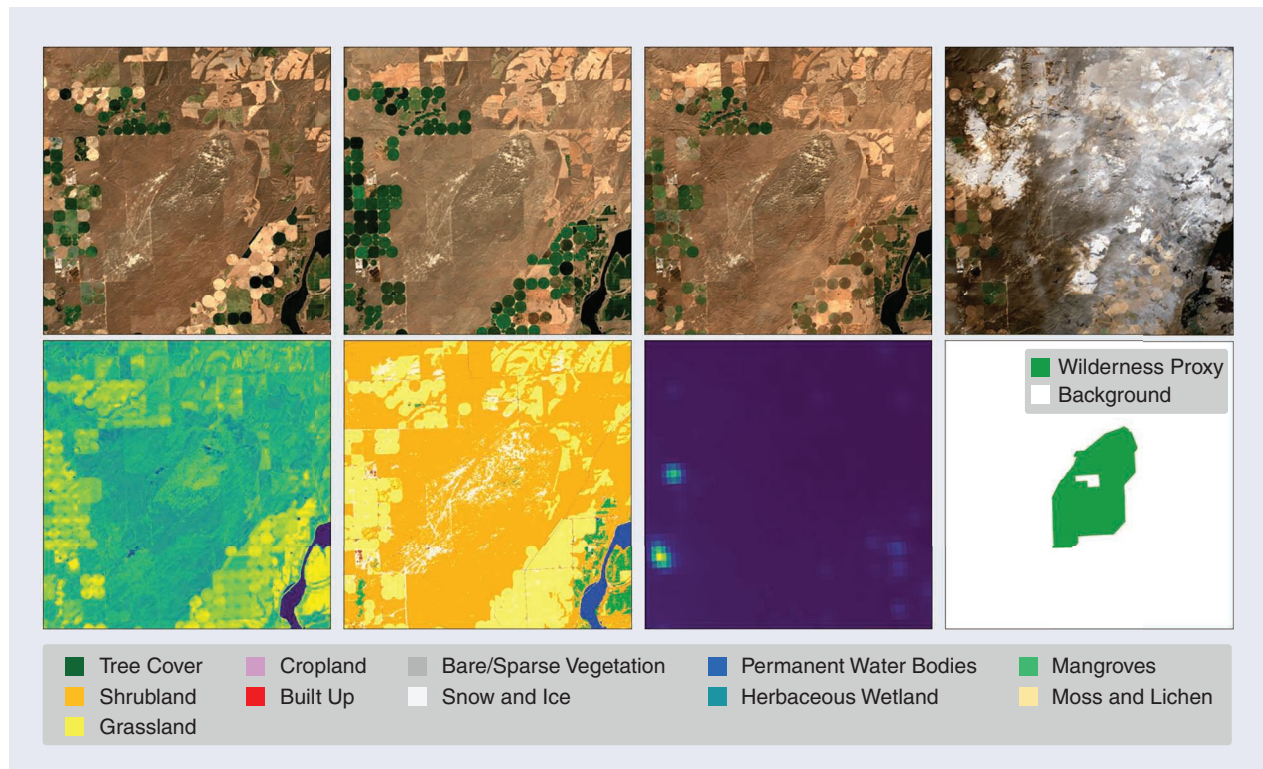


FIGURE 1. A wilderness sample from the MapInWild dataset (Juniper Dunes, United States, WDPA ID: 555556115). The first row shows the Sentinel-2 images of four seasons, from left to right: spring, summer, autumn, and winter. Second row, same order: Sentinel-1 image, ESA WorldCover map, VIIRS Nighttime Day/Night band, and WDPA annotation. For visualization purposes, the VV and VH bands of the Sentinel-1 image are treated as red and green bands, and only visible bands (red: B4, green: B3, and blue: B2) in the Sentinel-2 image are used. The ESA WorldCover map legend is given at the bottom the figure.

wide swath mode. The Level-2A Sentinel-2 image contains 10 spectral bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12) and is subject to mosaicking to mitigate the effect

caused by clouds. We stack the imagery of any area of interest (AOI) for a defined period of time in the year 2020, and for each band in each pixel, we select the value defining the 25% percentile regarding reflectivities observed in that time period. We empirically found that this way of compositing results in a more informative image than median compositing. To further widen the scope

of the dataset, we provide four seasons of Sentinel-2 images for every AOI. By adding a multitemporal value to the dataset by including all seasons for every AOI, we aim to further facilitate the investigation of wilderness mapping in a time-series manner. Further, rather than defining a single season period when querying the seasons, we take hemispheres of the AOI into account and define the season periods accordingly (i.e., Northern Hemisphere and Southern Hemisphere). We use the remaining sources (ESA WorldCover and the VIIRS) in their original forms and leave the end users with more flexibility. We performed reprojection and scaling to have the corresponding pixels across geodata sources in the same scale and projection, which uses the rubber-sheet algorithm as a registration method. The curation strategy is illustrated in Figure 2.

SAMPLING STRATEGY AND EXPORTING

Considering the difficulty of adequately sampling Earth's surface in terms of representability, we develop a guided

sampling approach taking the hard constraints of the WDPA polygons into account. Prior to filtering the images, sampling of the WDPA polygons takes place. Merely using the original WDPA polygons is not ideal because of the uneven distribution across continents and disparity in polygons (in terms of both volume and size). To this end, to ensure the versatility of the dataset and the spatial coverage of its samples, we design a climate map and land cover type-aware semiautomated sampling approach, where a subsampling of the polygons is performed while improving the spatial coverage of the WDPA polygons. The developed approach is illustrated in Figure 3. We start the process by applying a class-wise minimum area threshold (using the asset feature *GIS_AREA*) of 5 km² to category Ia and category Ib and 100 km² to category II. Then, after filtering out the nonterritorial areas, the sampling operation takes place, where we use a GEE function (namely, *StratifiedSample()*) that allows users to supervise the sampling process with weights that can be obtained from auxiliary data. For our case, we use the Köppen–Geiger climate classification map and the ESA WorldCover map as a proxy to maintain guidance over the sampling process. The sampling weights used in the *StratifiedSample()* function are formed by calculating the class-wise correlation between the WorldCover and climate classification maps (i.e., the frequency of each land cover class in each climate classification map). Consequently, the calculated weights are inversely normalized, and row-wise summations are used as sampling weights. Hence, oversampling of underrepresented polygons (and vice versa) is performed by following the observation that the representability of protected areas could be promoted by taking land cover and climate zones into account. Further, the polygons within a certain proximity (30 km² for category Ia and category Ib and 50 km² for category II) are removed. Ultimately, the resulting samples are exported for further analysis after fitting a bounding box with a size of 20 km × 20 km to the center of each polygon.

After exporting, the images are renamed after the WDPA polygon they contain and cropped into 1,920 × 1,920 pixel patches to remove the reprojection effect occurring near borders. We further add 108 AOIs from populous areas that exhibit human disturbance on Earth in various forms to strengthen the versatility of the dataset. The final distribution of the AOIs sampled from the WDPA polygons is given in Figure 4. The motivation behind this manual intervention, which increased the number of AOIs in the dataset from 910 to the final number of 1,018, is due to the potentiality that the model might overfit the wilderness areas in the form of forested areas (see the ESA WorldCover class distribution in Figure 5). A naming convention of 9000000XXX (where X is the identification number of the patch) is adopted when including the manually selected AOIs in the dataset. Overall, our dataset has 8,144 images (1,018 × 8, where 1,018 is the number of AOIs in the dataset and 8 is the number of the data sources each AOI contains)

WE USE THE KÖPPEN–GEIGER CLIMATE CLASSIFICATION MAP AND THE ESA WORLDCOVER MAP AS A PROXY TO MAINTAIN GUIDANCE OVER THE SAMPLING PROCESS.

TABLE 1. IUCN PROTECTED AREA CATEGORIES.

WDPA CLASS	DESCRIPTION
Ia: Strict nature reserve	Strictly protected areas where human presence is strictly limited and controlled
Ib: Wilderness area	Slightly modified areas with little human presence in the form of indigenous and local communities
II: National park	Functioning ecosystems subject to tourism through zoning
III: Natural monument or feature	Areas where preserving a particular feature hosting a cultural value is the dominant goal
IV: Habitat/species management area	Areas where flora species, fauna species, or habitats are aimed to be preserved and/or restored through informed interventions
V: Protected landscape/seascape	Areas where a distinct value is created by human presence over time
VI: Protected area with sustainable use of natural resources	Areas where the conservation of natural ecosystems and ecological processes take place

with a shape of 1,920 × 1,920 pixels. A sample image from the dataset is given in Figure 1.

QUALITY ASSESSMENT AND STATISTICS

With the intention of further improving the usability of the dataset, we conduct a quality assessment of the Sentinel-2 seasons. First, an evaluation platform is prepared and shared with many RS experts. The participants are then asked to annotate the informativeness of each Sentinel-2 season. Next, by performing a statistical analysis that takes the percentage of the votes, we 1) discard the patches with a discard score of higher than 50%, 2) assign a quality score between 0 and 10 to each season, and 3) calculate

the single temporal subset of each season using the quality scores calculated in the previous step. In cases where several seasons are assigned with the same quality score, we favor the season summer over the other seasons. In addition to the single temporal subsets, we also make the quality scores publicly available, which can be used to form case-specific single temporal subsets.

With this quality assessment, we not only refine the dataset and remove noninformative patches, but also boost the versatility of the dataset by generating quality scores and single temporal subset seasons assigned to each season and patch, respectively. We advise the researchers to make use of the single temporal subset seasons defined for each

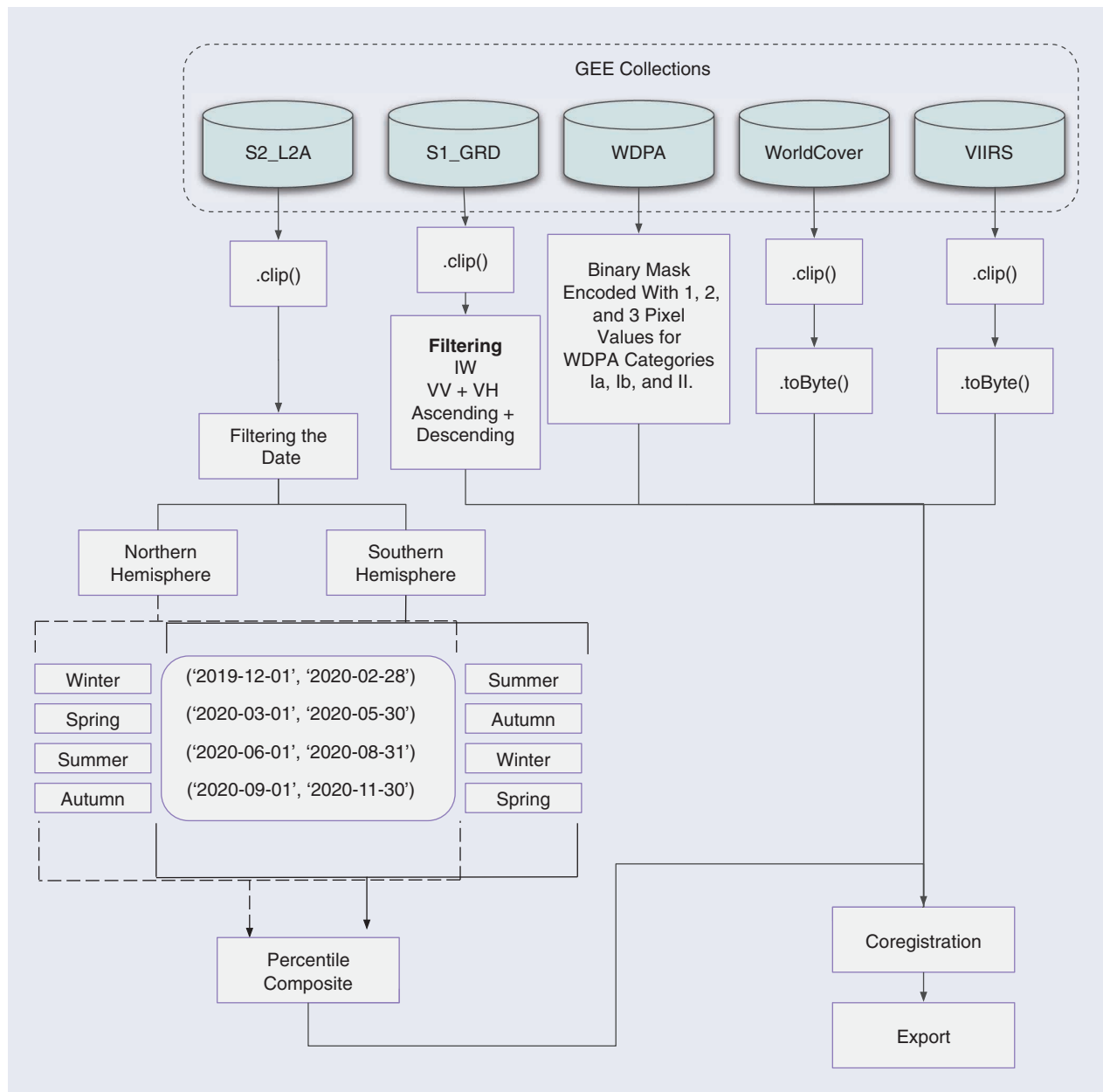


FIGURE 2. Curation workflow: querying and filtering the geodata sources in the MapInWild dataset. L2A: Level-2A; GRD: ground range distance.

patch should they be solely interested in the single temporal information.

The final distribution of the AOIs sampled from the WDPA polygons is given in Figure 4. Figure 5 shows the percentage of ESA WorldCover classes falling in the AOIs.

BASELINE EXPERIMENTS

In the following experiments, we aim to discover the concept of wilderness in two different ways: by semantic segmentation and scene classification with a focus on sensitivity analysis. Both experiments work toward providing initial efforts in understanding wilderness. For both experiments, out

of 1,018 patches, we used 80%, 10%, and 10% of them for training, validation, and test sets, respectively. The patch IDs of the sets are made publicly available, and the distribution of the sets is shown in Figure 4. For the semantic segmentation experiment, we unify the three protected area classes in the annotation sources to form binary wilderness masks.

SEMANTIC SEGMENTATION

EXPERIMENTAL SETUP

For this experiment, we adopt the U-Net architecture [33] with ResNeSt 14d [34]. The model consists of approximately

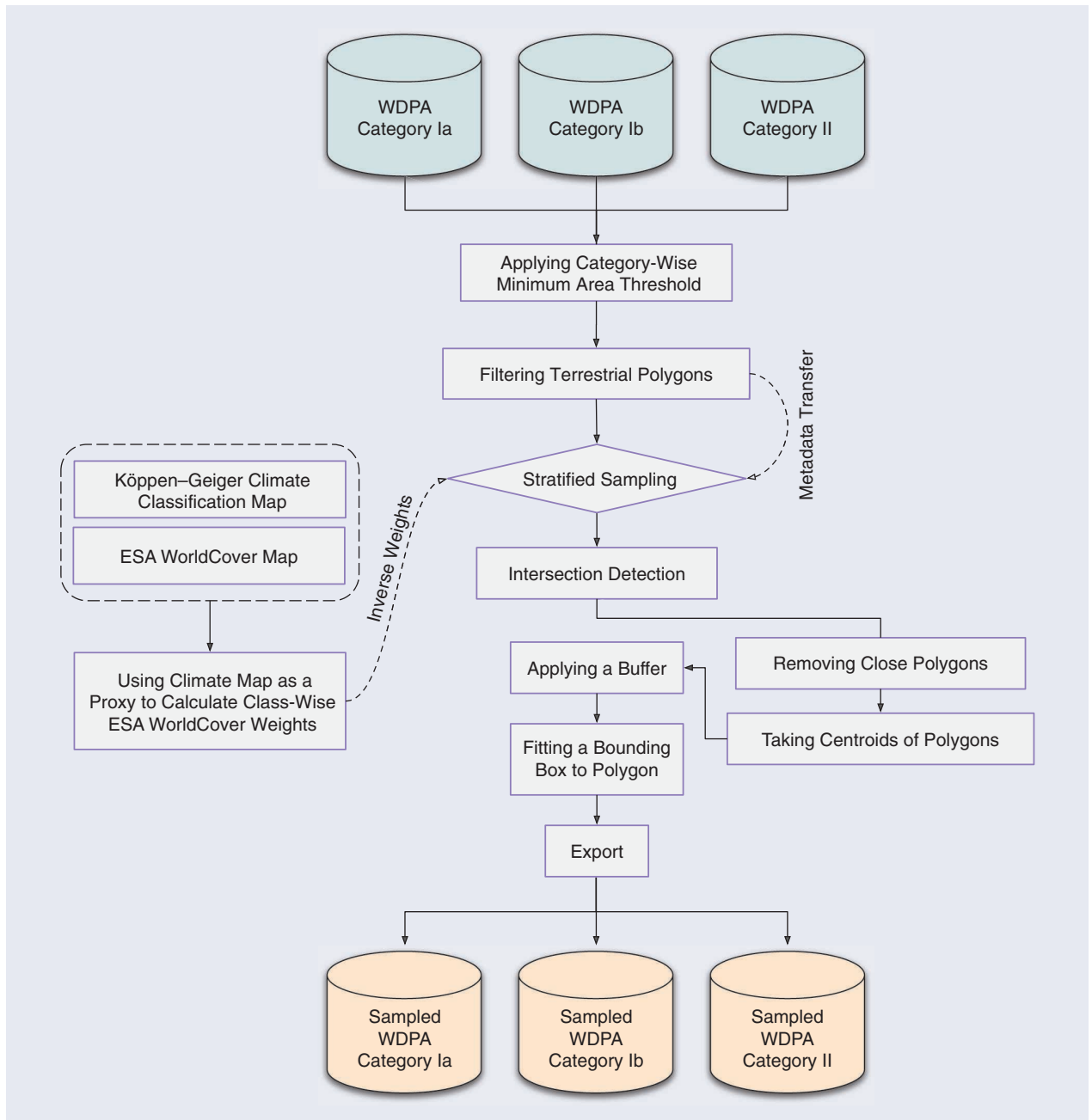


FIGURE 3. Sampling workflow: the guided stratified sampling of the WDPA polygons. The WDPA polygons that are varying in size and shape are filtered through a semiautomated sampled process that is guided by the climate zones and the land cover types.

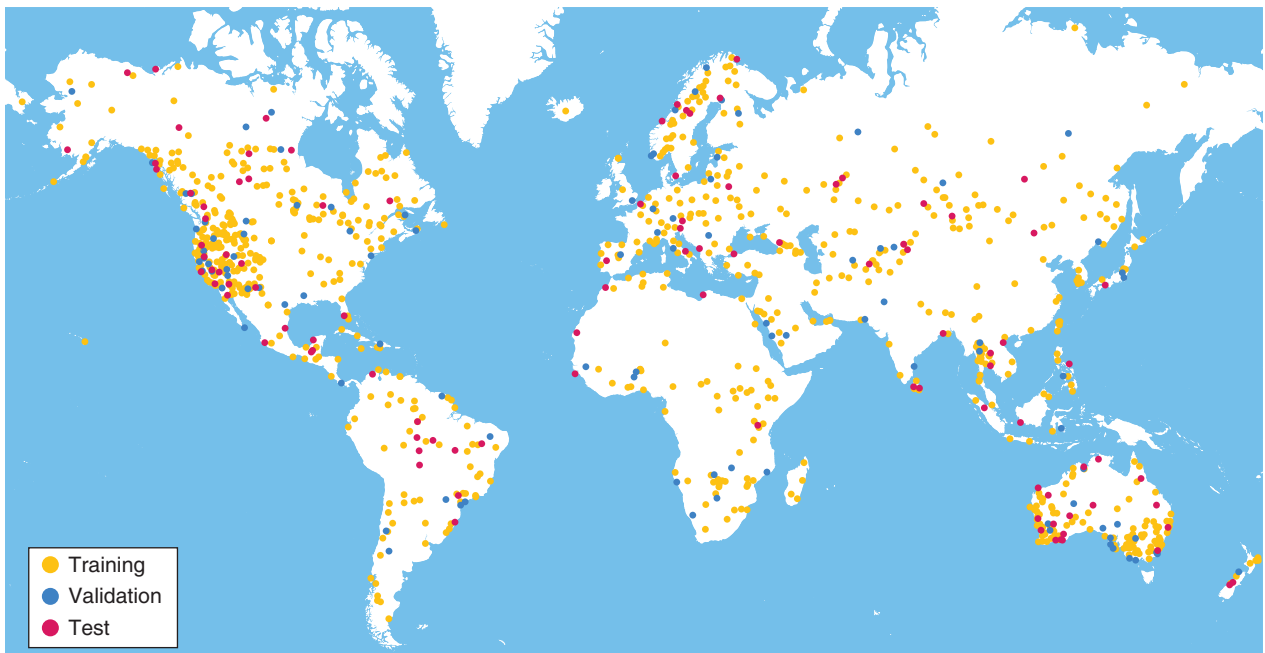


FIGURE 4. Distribution of the AOIs and training, validation, and test sets across continents. The points represent the center of each study area and are enlarged for better visualization.

8 million parameters. As an optimization algorithm, we use Adam with beta values of (0.9, 0.999) and an epsilon value of 10^{-8} . We select the learning rate as 0.0001 and adopt the cosine annealing learning rate schedule with the maximum number of iterations set to 10. We use the dice loss to quantify the cost value and set the batch size as 32. During the training process (e.g., on the fly), we randomly extract patches with a shape of 512×512 pixels from the input images and feed them to the model. We use the three-band (B4, B3, and B2) Sentinel-2 image queried with the single temporal subset seasons and utilize ImageNet pretrained weights during the feature extraction phase. Furthermore, we configure mixed-precision training to shorten the training and inference time and decrease the required amount of memory. We train the model on an Nvidia Tesla V100 SCM3 32GB. The training process is terminated with an early stopping callback with a patience (the number of epochs without improvement in the metric observed) of 8. The achieved overall pixel-wise classification accuracy, intersection over union (IoU) score, and F1 score are shown in Table 2. The quantitative results are shown in Figure 6.

SCENE CLASSIFICATION AND SENSITIVITY ANALYSIS

In this experiment, we perform scene classification with a subsequent sensitivity analysis. To this end, we use the single temporal subset of the Sentinel-2 data and extract tiles with a height and width of 256 pixels each. The tiles are extracted in such a way that they are located exclusively either within or without the WDPA regions without overlapping. Thus, we get 24,101 tiles within WDPA regions

(label 1) and 17,795 tiles outside of WDPA regions (label 0). The proportions of the training, validation, and test set remain approximately the same.

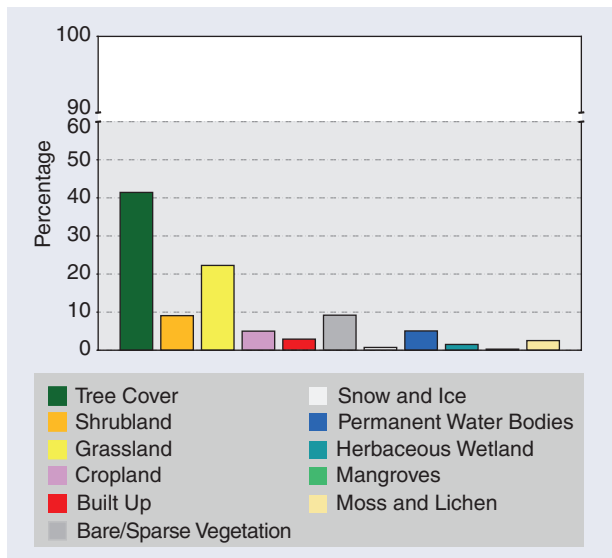


FIGURE 5. The pixel-level breakdown of final AOIs in terms of land cover. The land cover information is obtained from the ESA World-Cover map existing in the dataset.

TABLE 2. QUALITATIVE RESULTS (%): BASELINE MODEL FOR THE SEMANTIC SEGMENTATION TASK.

MODEL	BACKBONE	IOU	ACCURACY	F1
U-Net	ResNeSt 14d	69.1	76.31	81.73

We perform ASOS according to [21] to create high-resolution sensitivity maps of our test samples to assess whether specific regions are sensitive toward wilderness or nonwilderness. The authors provide neural network architecture, a specific training procedure, and the approach for sensitivity analysis. We proceed in the same way and refer to their work for a detailed description of the methodology.

The neural network consists of a modified form of the U-Net of [33] and a simple classifier network, shown in Figure 7. The activation map at the interface of these two networks has three channels and is the same size as the input image. The model has about 2 million parameters.

CLASSIFICATION

We run the same training procedure as in [21] and use a maximum learning rate of 0.01, a weight decay of 0.0001, and a batch size of 32. We train the model for 50 epochs on an Nvidia Quadro RTX 4000 (8 GB GDDR6) for about 6 h.

The achieved overall classification accuracies are 92% for the training set, 82% for the validation set, and 74% for the test set. The confusion matrix of the test set is shown in Table 3.

SENSITIVITY ANALYSIS

After the model has been trained, we run the ASOS analysis on all correctly classified training samples. For this, specific pixels are occluded in the activation maps by setting them to zero.

This changes the classification scores, and the deviations are a measure of the sensitivities of the occluded parts. The occlusions are determined from the activation space in which each of the three axes corresponds to one of the three channels of the activation maps. Pixels that are close to each other in the activation space are occluded simultaneously. The resulting sensitivities are visualized in Figure 8. Low-density areas are not included in this mapping. Knowing the sensitivities, we can predict sensitivity maps for any input image. A selection of the test samples is shown in Figure 9.

DISCUSSION

With the semantic segmentation experiment, we study the concept of wilderness on a pixel scale. The model used in this experiment outputs densely predicted segmentation maps for a given test image with a shape of $1,920 \times 1,920$, as shown in Figure 6. The first sample shows parts of the Rio Novo National Park, where the park is seen as divided in two by a river. We see that the model predicts the whole patch (except for the deforestation areas) as wilderness, while the WDPA proxy contains annotation only on the right side of the river. The second sample shows the Phu Pha Man District in Thailand. Here, the proxy loosely aligns with the models' output. The output segmentation map here filters out the areas with a human influence (i.e., a suburban area surrounded by cultivated lands). The third sample is located at Santa Bárbara Ecological



FIGURE 6. Baseline results for the test images with the IDs 351819, 39516, 19442, and 900000068. Row-wise, from left to right: input image, wilderness proxy, and model's output. Input images are the single temporal subsets of the Sentinel-2 images for the corresponding IDs.

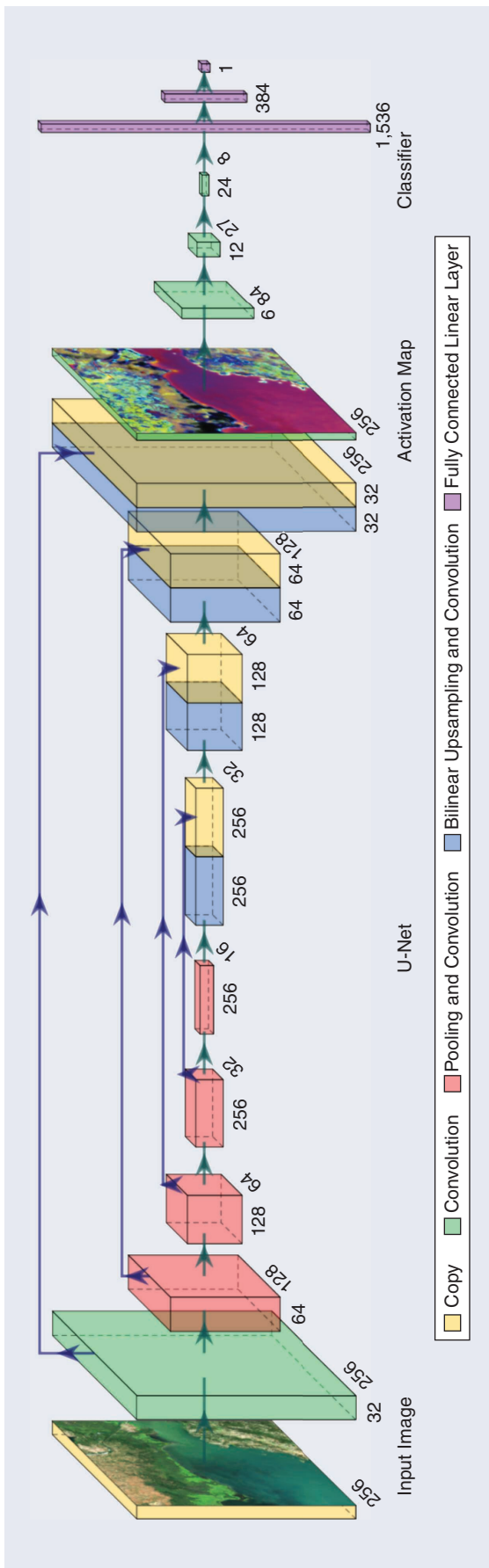


FIGURE 7. For the sensitivity analysis, we use a neural network consisting of a modified U-Net and an image classifier. The activation map at the interface is essential for determining the sensitivities. (Source: [21]; adapted slightly.)

Station in Brazil. Here, the segmentation map does not include any annotation. This could be explained by the observation that this area exhibits human influence in the form of roads and arable lands, which clearly does not comply with the “undisturbedness” and “undevelopedness” states of the wilderness areas. The last sample shows Cape Coral, Florida. This is one of the samples we have manually added to the dataset (notice the 9 at the beginning of the ID). This sample shows that the model does not annotate an urban area as wilderness, although there are golf courses, inner city parks, football courts, and forest areas on the perimeter. This sample also set forth the ability of the model to capture the “disturbedness” and “developedness” of an area. All things considered, this experiment exhibits that the model is competent to discover wilderness areas on a pixel scale. While doing so, the model provides a more viable annotation of wilderness areas as an output. With this observation, we go a step further and analyze the models’ output with the use of explainable ML techniques to take a deeper look at the models’ decisions.

TABLE 3. THE CONFUSION MATRIX OF THE TEST DATASET FOR THE CLASSIFICATION TASK. PERCENTAGES ARE GIVEN IN BRACKETS.

LABEL	PREDICTION	
	NONWILD	WILD
Nonwild	1,062 [62.65%]	633 [37.34%]
Wild	439 [17.79%]	2,028 [82.2%]

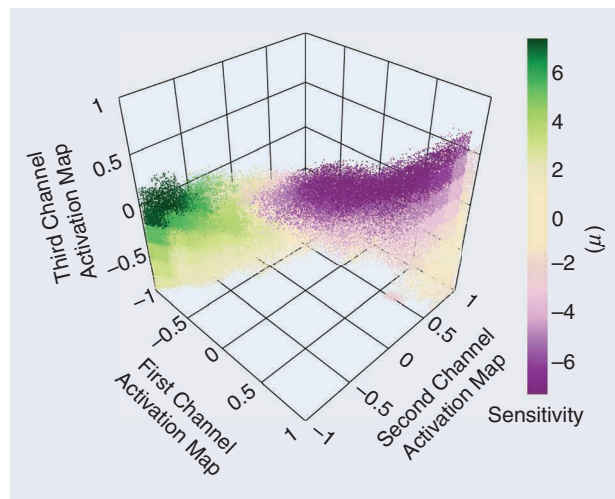


FIGURE 8. The activation space of the sensitivity analysis. Each axis corresponds to the values of the respective channel in the activation maps. The greener the vectors, the more they are sensitive to wilderness characteristics; the more violet, the more sensitive to nonwilderness characteristics. Low-density areas are not included in the sensitivity mapping, and the corresponding vectors are not shown here.

Applying ASOS to the MapInWild dataset, we demonstrate an example task for explainable ML. Training a neural network on classification, we can predict the high-resolution sensitivity maps shown in Figure 9. The first sample shows parts of the Australian national park Karijini. It is centrally split by a mine and a railway. The human influence on the surroundings appears to be so small that the model does not highlight them as such. The second sample is located at the Sid's Mountain Wilderness Study Area in Utah, United States. The model detects some parts that are mainly outside the study area as nonwild. The inner parts of the study area seem to have wild characteristics. The third sample shows the Skadar Lake National Park in Montenegro and Albania. At the

top, agricultural fields reach inside the national park; this is detected by the model. The lake itself has mainly not been predicted because of high uncertainty. This goes with the MapInWild dataset, which has been created by specifically looking for nonwater polygons. The fourth sample shows the small Skwaha Lake Ecological Reserve in Canada. This reserve has, according to the model, mixed characteristics. The two regions detected as nonwild are valleys. The left one includes several streets, and the right region includes Trans-Canada Highway 1 along a river.

We provide the preceding experiments as an initial effort in addressing the novel task of wilderness mapping on the large-scale benchmark dataset MapInWild. The

task has first been investigated on a pixel scale within the frame of semantic segmentation, where the imperfect wilderness annotations have been used as a proxy when mapping the wilderness areas. On the one hand, there are cases where the learner extrapolates the wilderness area to the entire scene as the wilderness area in the annotation is of similar appearance to the remaining area, which is annotated as nonwilderness (the first sample in Figure 6). This behavior of the learner exhibits its ability to learn the concept of wilderness from imperfect annotations. On the other hand, it is seen that, when mapping the wilderness areas, the learner dampens the presence of wilderness characteristics shown in the proxy. Although there is an annotated wilderness characteristic in the center of the area (the third sample in Figure 6), the existing human influence in the surroundings might be negatively affecting the essence and form of the wilderness concept realized and discovered by the learner. The concept realized might contain a learned pattern on 1) the size of a wilderness area, 2) the distance to the nearest area under human influence, and 3) characteristics of the site in terms of land use and land cover. Similar to this behavior, the learner shows strength in leaving out the resemblance of inner city parks and forests with a wilderness area in the form of a forested area (the last sample of Figure 6). In the middle of the learner's behavior



FIGURE 9. Baseline results for the test images with IDs 64113, 374681, 16385, and 18415. Row-wise, from left to right: input image, wilderness proxy, and sensitivity map. The color scale of the sensitivity maps is given in Figure 8. Areas with a low density in the activation space are not mapped to sensitivities and are colored gray in the sensitivity maps.

scale, there lies a typical representation of a semantic segmentation task in which the learner approximates the given annotation (the second sample in Figure 6). While doing so, the learner struggles with the borders of the given wilderness area, which can be explained by 1) the edge-effect [31], which complicates the segregation of ecological units from the previous perspective, and 2) the imperfect annotations used when training the learner.

Later, the learner has been further investigated by employing an interpretable-by-design architecture to study the patterns in the decision-making process. The high-resolution sensitivity maps in Figure 9 make it evident that the learner holds a deeper understanding of wilderness that disentangles wilderness from human influence. The maps produced in this experiment provide pixel-level sensitivity information that could be utilized in the process of inferring new scientific insights.

Although the behavior of the learner can often be explained away with some confidence, in the light of the experimental results explained previously, it is still unclear when and why the learner behaves in certain ways. Motivated by this observation, we urge environmental science, conservation, computer science, and RS researchers to study the ambiguity in the ill-defined elements of nature to better monitor, understand, and protect nature, our home.

SUMMARY AND CONCLUSION

With this article, we 1) introduce a novel task of wilderness mapping and 2) publish MapInWild, a large-scale benchmark dataset curated for the task of wilderness mapping. MapInWild is a multimodal dataset and comprises various geodata acquired and formed from a diverse set of RS sensors. The dataset consists of 8,144 images with a shape of $1,920 \times 1,920$ pixels and is approximately 350 GB in size. The images are weakly annotated with three classes derived from the WDPA: strict nature reserves, wilderness areas, and national parks. With MapInWild, for the purpose of deepening our understanding of what makes nature wild, we embark on the complications induced by the ambiguity of the term *wilderness* and study the vagueness in nature and propose our dataset as a test bed for ML research concerning environmental RS. We are convinced that getting closer to understanding the concept of wilderness is of great value to the community to further bridge the gap between DL applied to environmental RS and conservation. Both the MapInWild dataset and the code are publicly available at <https://dataverse.harvard.edu/dataverse/mapinwild> and <https://github.com/burakekim/MapInWild>.

ACKNOWLEDGMENT

This work was supported by the German Research Foundation (DFG project MapInWild, Grant RO 4839/5-1/SCHM 3322/4-1). The authors acknowledge the computing time

granted by the Institute for Distributed Intelligent Systems and provided on the GPU cluster Monacum One at the University of the Bundeswehr Munich.

AUTHOR INFORMATION

Burak Ekim (burak.ekim@unibw.de) is with the Institute of Space Technology and Space Applications, Department of Aerospace Engineering, University of the Bundeswehr Munich, 85577 Neubiberg, Germany.

Timo T. Stomberg (timo.stomberg@uni-bonn.de) is with the Institute of Geodesy and Geoinformation, University of Bonn, 53115 Bonn, Germany.

Ribana Roscher (ribana.roscher@uni-bonn.de) is with the Institute of Geodesy and Geoinformation, University of Bonn, 53115 Bonn, Germany.

Michael Schmitt (michael.schmitt@unibw.de) is with the Institute of Space Technology and Space Applications, Department of Aerospace Engineering, University of the Bundeswehr Munich, 85577 Neubiberg, Germany.

REFERENCES

- [1] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.
- [2] M. Schmitt, S. A. Ahmadi, and R. Hänsch, "There is no data like more data - Current status of machine learning datasets in remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, Jul. 11, 2021, pp. 1206–1209, doi: 10.1109/igarss47720.2021.9555129.
- [3] G. Sumbul et al., "BigEarthNet-MM: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [Software and Data Sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 174–180, Sep. 2021, doi: 10.1109/MGRS.2021.3089174.
- [4] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS: A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," in *Proc. ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, Sep. 2019, vol. IV-2/W7, pp. 153–160, doi: 10.5194/isprs-annals-IV-2-W7-153-2019.
- [5] X. X. Zhu et al., "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [Software and Data Sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 3, pp. 76–89, Sep. 2020, doi: 10.1109/MGRS.2020.2964708.
- [6] X. Sun et al., "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 184, pp. 116–130, Feb. 2022, doi: 10.1016/j.isprsjprs.2021.12.004.
- [7] D. Silvestro, S. Gorla, T. Sterner, and A. Antonelli, "Improving biodiversity protection through artificial intelligence," *Nature Sustain.*, vol. 5, no. 5, pp. 415–424, Mar. 24, 2022, doi: 10.1038/s41893-022-00851-6.
- [8] C. Persello et al., "Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities," *IEEE Geosci. Remote*

- Sens. Mag.*, vol. 10, no. 2, pp. 2–30, Jun. 2022, doi: 10.1109/mgrs.2021.3136100.
- [9] D. Tuia et al., “Perspectives in machine learning for wildlife conservation,” *Nature Commun.*, vol. 13, no. 1, Feb. 9, 2022, Art. no. 792, doi: 10.1038/s41467-022-27980-y.
- [10] B. Ekim and M. Schmitt, “Mapinwild: A dataset for global wilderness mapping,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 231–234, doi: 10.1109/IGARSS46834.2022.9883217.
- [11] J. Song, S. Gao, Y. Zhu, and C. Ma, “A survey of remote sensing image classification based on CNNs,” *Big Earth Data*, vol. 3, no. 3, pp. 232–254, 2019, doi: 10.1080/20964471.2019.1657720.
- [12] 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, May 2018, doi: 10.1109/TGRS.2017.2783902.
- [13] C. Qiu, L. Liebel, L. H. Hughes, M. Schmitt, M. Korner, and X. X. Zhu, “Multitask learning for human settlement extent regression and local climate zone classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2020.3037246.
- [14] G. Cheng, X. Sun, K. Li, L. Guo, and J. Han, “Perturbation-seeking generative adversarial networks: A defense framework for remote sensing image scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022, doi: 10.1109/TGRS.2021.3081421.
- [15] N. Lang, N. Kalischek, J. Armston, K. Schindler, R. Dubayah, and J. D. Wegner, “Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles,” *Remote Sens. Environ.*, vol. 268, Jan. 2022, Art. no. 112760, doi: 10.1016/j.rse.2021.112760.
- [16] Y. Xu, B. Du, and L. Zhang, “Self-attention context network: Addressing the threat of adversarial attacks for hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 30, pp. 8671–8685, Oct. 2021, doi: 10.1109/TIP.2021.3118977.
- [17] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833, doi: 10.1007/978-3-319-10590-1_53.
- [18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [19] M. T. Ribeiro, S. Singh, and C. Guestrin, “why should I trust you?: Explaining the predictions of any classifier” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
- [20] K. Bastmeijer, *Wilderness Protection in Europe: The Role of International, European and National Law*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
- [21] T. Stomberg, T. Stone, J. Leonhardt, I. Weber, and R. Roscher, “Exploring wilderness characteristics using explainable machine learning in satellite imagery,” 2022, *arXiv:2203.00379*.
- [22] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, “Explainable machine learning for scientific insights and discoveries,” *IEEE Access*, vol. 8, pp. 42,200–42,216, Feb. 2020, doi: 10.1109/ACCESS.2020.2976199.
- [23] L. Hannah, J. L. Carr, and A. Lankerani, “Human disturbance and natural habitat: A biome level analysis of a global data set,” *Biodiversity Conservation*, vol. 4, no. 2, pp. 128–155, 1995, doi: 10.1007/BF00137781.
- [24] B. Ekim, Z. Dong, D. Rashkovetsky, and M. Schmitt, “The naturalness index for the identification of natural areas on regional scale,” *Int. J. Appl. Earth Observ. Geoinformation*, vol. 105, Dec. 2021, Art. no. 102622, doi: 10.1016/j.jag.2021.102622.
- [25] J. R. Allan, O. Venter, and J. E. M. Watson, “Temporally inter-comparable maps of terrestrial wilderness and the Last of the Wild,” *Scientific Data*, vol. 4, no. 1, pp. 1–8, 2017, doi: 10.1038/sdata.2017.187.
- [26] E. W. Sanderson, M. Jaiteh, M. A. Levy, K. H. Redford, A. V. Wambenebo, and G. Woolmer, “The human footprint and the last of the wild: The human footprint is a global map of human influence on the land surface, which suggests that human beings are stewards of nature, whether we like it or not,” *BioScience*, vol. 52, no. 10, pp. 891–904, 2002, doi: 10.1641/0006-3568(2002)052[0891:TH FATL]2.0.CO;2.
- [27] T. Stomberg, I. Weber, M. Schmitt, and R. Roscher, “Jungle-Net: Using explainable machine learning to gain new insights into the appearance of wilderness in satellite imagery,” in *Proc. ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, 2021, vol. V-3-2021, pp. 317–324, doi: 10.5194/isprs-annals-V-3-2021-317-2021.
- [28] “Protected planet: The World Database on Protected Areas (WDPA).” UNEP-WCMC and IUCN. Accessed: Apr. 25, 2022. [Online]. Available: <https://www.iucn.org/theme/protected-areas/our-work/world-database-protected-areas>
- [29] “Guidelines on wilderness in natura 2000: Management of terrestrial wilderness and wild areas within the natura 2000 network,” European Commission, Brussels, Belgium, Tech. Rep. 069, 2013.
- [30] B. G. Mackey, R. G. Lesslie, D. B. Lindenmayer, and H. A. Nix, “Wilderness and its place in nature conservation in Australia,” *Pacific Conservation Biol.*, vol. 4, no. 3, p. 182, 1998, doi: 10.1071/PC980185.
- [31] C. Murcia, “Edge effects in fragmented forests: Implications for conservation,” *Trends Ecol. Evol.*, vol. 10, no. 2, pp. 58–62, 1995, doi: 10.1016/S0169-5347(00)88977-6.
- [32] 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/
- [33] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer International Publishing, 2015, pp. 234–241.
- [34] H. Zhang et al., “ResNeSt: Split-attention networks,” 2020. [Online]. Available: <https://arxiv.org/abs/2004.08955>