Gaining access to labeled reference data is one of the great challenges in supervised machine-learning endeavors. This is especially true for an automated analysis of remote-sensing images on a global scale, which enables us to address global challenges, such as urbanization and climate change, using state-of-the-art machine-learning techniques. To meet these pressing needs, especially in urban research, we provide open access to a valuable benchmark data set, So2Sat LCZ42, which consists of local-climate-zone (LCZ) labels of approximately half a million Sentinel-1 and Sentinel-2 image patches in 42 urban agglomerations (plus 10 additional smaller areas) across the globe.

This data set was labeled by a group of domain experts following a carefully designed workflow and evaluation process. We conducted rigorous quality assessment with independent label voting by domain experts, which is rarely done in other labeled remote sensing data sets. The data set achieved an overall confidence of 85%. We believe this is a first step toward an unbiased, globally distributed data set for urban growth monitoring using machine-learning methods, because LCZ analysis provides a rather objective measure compared with many other semantic land-use and land-cover classification systems. It measures morphology, compactness, and height of urban areas, which are less dependent on human activity and culture. Furthermore, such large-scale labeled data with uncertainty measures can serve as a benchmark for cutting-edge machine-learning research specific to Earth observation (EO), such as automatic topology learning, data fusion, modeling uncertainties in machine learning, and many more. This data set can be accessed at http://doi.org/10.14459/2018mp1483140.

THE NEED FOR LOCAL-CLIMATE-ZONE LABELS
The production of land-use/land-cover (LULC) maps at large or even global scales is an essential task in the field of remote sensing. These maps can provide valuable input for many societal questions, such as understanding human poverty or climate change, supporting the conservation of biodiversity and ecosystems, and providing stakeholder information for disaster management and sustainable urban development [1].

Urbanization is undoubtedly the most important megatrend in the 21st century after climate change. Currently, half of humanity—3.5 billion people—lives in cities. Shockingly, one billion of them still live in slums. Therefore, sustainable urban development has become one of the 17 sustainable development goals of the United Nations. Today, sustainable development increasingly depends on the successful management of urban growth, especially in developing countries, where the pace of urbanization is projected to be the fastest, according to World Urbanization Prospects: The 2018 Revision [2]. LULC maps enable us to describe, track, and manage urban growth in an objective and consistent manner.

Examples of global LULC products created by the remote sensing community include the Global Urban Footprint [3], [4], produced from synthetic aperture radar (SAR) data acquired by the TanDEM-X mission; the Global Human Settlement Layer, created from global, multitemporal archives of fine-scale satellite imagery, census data, and volunteered geographic information [5]; and the Finer Resolution Observation and Monitoring of Global Land Cover and Globe-Land30 data sets, generated from 30-m-resolution Landsat data [6]. This list is not exhaustive; however, these products all provide semantic labels of urban and nonurban or even more specific classes. These semantic labels are often subjective (to human interpretation) and culture-dependent. For example, the definitions of urban and nonurban areas might be drastically different in Europe and Africa and from person to person.

LOCAL CLIMATE ZONES IN GLOBAL URBAN MAPPING
For consistent analysis across the globe, an objective and culture-independent classification scheme for urban areas
is pressingly needed. After extensive research, we turned to LCZs, which were originally developed for metadata communication of observational urban heat island studies [7]. There are 17 classes in the LCZ classification scheme; 10 are built classes and seven are natural classes. They are based on climate-relevant surface properties on the local scale, which are mainly related to 3D surface structures (e.g., the height and density of buildings and trees), surface cover (e.g., vegetation or paving), and anthropogenic parameters (such as human-based heat output).

A schematic drawing of the 17 classes is shown in Figure 1(a). The 10 urban classes describe the morphology of the area, including the density and height of the buildings as well as the percentage of impervious surface. The urban classes are mostly shown in red, with decreasing intensities as the building density and height decrease from compact high-rise to open low-rise. Figure 1(b) shows the LCZ classification of Vancouver, Canada, created by the authors. The dark red part marked by the yellow rectangle is downtown Vancouver, where most of Vancouver’s high-rise buildings are located. The light red part of the classification map is mostly low-rise residential houses. As a reference, the Google image of this area in Figure 1(c).

Because the LCZ classes are defined by their physical properties, they are generic and applicable to cities across the world, offering the potential to compare different areas of various cities with trenchant distinctions representing the heterogeneous thermal behavior within an urban environment [8]. In addition to the increasing impact on worldwide climatological studies, such as the cooling effect of green infrastructure and microclimatic effects on town peripheries [9]–[18], researchers have recently started to use the LCZ approach to classify the internal structure of urban areas, providing promising information for various applications, such as infrastructure planning, disaster mitigation, health and green-space planning, and population assessment [19], [20]. The remote sensing community also addressed this topic by organizing the 2017 IEEE Geoscience and Remote Sensing Society Data Fusion Contest, which had the goal of LCZ classification [21].

RELATED WORK IN LOCAL-CLIMATE-ZONE CLASSIFICATION

COMMUNITY-BASED LOCAL-CLIMATE-ZONE MAPPING

A significant part of the existing development of LCZ classification is community-based, large-scale LCZ mapping using freely available Landsat data and software [23]–[25]. The World Urban Database and Portal (WUDAPT) [22], a community-driven initiative, was organized by researchers to map high-quality LCZ maps worldwide. Within this framework, currently, almost 100 cities worldwide have been mapped with moderate quality, providing sufficient detail for certain model applications [26]. LCZ maps of tens of cities, after undergoing quality assessment and generation of metadata, are now openly available at the WUDAPT portal. More recently, LCZs of Europe are being mapped as part of the WUDAPT project, with data from Sentinel-1, Sentinel-2, and the Defense Meteorological Program Operational Linescan System Nighttime Lights product [27].

These community-based efforts mark the first step toward a more synergetic cooperation among researchers. However, multiple studies have reported that the quality of the produced LCZ maps is inconsistent [28], [29], as the procedures strongly rely on the knowledge of individual volunteers. For example, the methods of community-based LCZ mapping mainly consist of 1) labeling of reference data in Google Earth and 2) classification using shallow learning algorithms, such as random forest (RF) in geographic information system software, a process detailed in [8].
ALGORITHMIC DEVELOPMENT

Therefore, significant development is still needed to achieve global LCZ mapping because of the lack of high-quality labels and transferable classifiers for worldwide deployment. There are various promising classifiers for LCZ recently proposed by different research groups, including RF, support vector machines (SVMs) [30], canonical correlation forests [31], rotation forests [21], gradient-boosting machines [33], and ensembles of multiple classifiers [34]. The data used are mainly satellite data in the optical and microwave ranges, such as Landsat, Sentinel-1, and Sentinel-2 images.

Recently, fusing of multisource data, such as satellite images and Google Street View, has also been investigated for LCZ classification [35]. Deep learning certainly played an important role in LULC using remote sensing data [36]. Multiple algorithms based on convolutional neural networks, such as residual neural network and ResNeXt, [35], [37]–[42] have been developed. These approaches provide satisfying results for specific areas. However, according to [8], [26], and [43], regional variations in vegetation and artificial materials as well as significant variations in cultural and physical environmental factors cause large intraclass variability of spectral signatures. One existing effort to further improve LCZ classification results is developing more robust machine-learning models with high generalization ability to facilitate efficient upscaling in a reasonable time frame [27], [43]. Deep-learning-based models have been shown to have better generalization ability; thus, they can be better exploited for LCZ classification [36], [40].

Despite active algorithmic development, the global transferability of a machine-learning LCZ model requires a large quantity of globally distributed and reliable reference data as a first step. Such a data set is nonexistent in the community; this task is addressed in this article.

CONTRIBUTIONS OF THIS ARTICLE

THE DATA SET

To answer the pressing need for LCZ training data sets, we carefully selected and labeled 42 urban agglomerations plus 10 additional smaller areas across all of the continents (except Antarctica) around the globe. Their geographic distribution can be seen in Figure 2. Many polygons in those cities were manually labeled by the authors. By projecting these labels to the corresponding coregistered Sentinel-1 and Sentinel-2 images, we obtained 400,673 pairs of corresponding Sentinel-1 SAR and Sentinel-2 multispectral image patches with LCZ labels. An impression of the Sentinel image patch pairs in the data set can be seen in Figure 3. However, the actual patches in the data set have a dimension of 320 × 320 m, which is smaller than the visualization in Figure 3. Accompanying this article, we provide open access to this high-quality So2Sat LCZ42 data set to the research community. This is meant to foster the development of fully automatic classification pipelines based on modern machine-learning approaches and support the accelerated use of LCZ mapping at global scale.

IMPROVED LABELING WORKFLOW

We found that merely following the definition of LCZs in [9] and the labeling process mentioned in WUDAPT is not optimal for a joint labeling activity by a group of people because of the vague definitions of some LCZ classes. To ensure the highest possible quality of the result, we designed a rigorous labeling workflow and decision rules, shown in Figure 4 and “Decision Rule of Local-Climate-Zone Labeling,” respectively. Meetings were conducted before and during the labeling process to calibrate our understanding of the definitions of the 17 classes. Afterward, the labeling
results from each member of the labeling crew were visually inspected by a different person to spot obvious errors. Finally, we conducted a quantitative evaluation of the label quality. The entire rigorous labeling processing took approximately 15 person-months.

RIGOROUS LABEL-QUALITY ASSESSMENT

Similar to any remote sensing product, reference labels must have error bars to indicate their trustworthiness, but such a quality measure rarely appears in data sets of remote sensing image labels. As mentioned previously, we conducted a rigorous quantitative evaluation of 10 cities in the data set by having a group of remote sensing experts cast 10 independent votes on each labeled polygon, to identify possible errors and assess human labeling accuracy. “Human confusion matrices” per polygon and per pixel were created, where the confidence of individual classes can be seen. In general, our human labels achieve 85% confidence. This confidence number can serve as a reference accuracy for the machine-learning models trained on this data set.

So2Sat LCZ42 DATA SET CREATION

A four-phase labeling process was designed to maximize label consistency and minimize human error, consisting of learning, labeling, visual validation, and quantitative validation phases (Figure 4). The detailed procedures for each phase are introduced in this section. We also prepared the corresponding Sentinel-1 and Sentinel-2 images of the 52 areas, and
proper preprocessing procedures were performed on the two types of images.

**CREATING THE LABELS**

**LEARNING PHASE**

The learning phase aims at creating a standard for team members who conduct the labeling (referred to as the labeling crew). The reasons are twofold. First, the definition of LCZ classes (given in [9] and listed in Table 1) are not mutually disjoint (e.g., class 3, compact low-rise, and class 8, large low-rise), and their union does not describe the entirety of Earth’s surface. That is, that some areas do not fall into any of the LCZ classes, and some can be categorized according to multiple classes. Second, interpretations of the definitions by different individuals still differ from each other.

The labeling crew started by building a visual impression of different LCZ classes by viewing aerial images on Google Earth and then moved toward a quantitative understanding of each class. As a result, we constructed a quantitative-labeling decision rule according to the literal definition. This is shown in Figure S1. An examination of the labeling learning course was conducted before actual labeling began, for which everyone in the labeling crew cast a vote on many selected scenes. Ambiguous scenes were selected and discussed to calibrate everyone’s understanding.

**LABELING PHASE**

The labeling phase followed a standard procedure defined in the WUDAPT project [22]. First, each member of the labeling crew claimed a few cities among the 52 cities and defined a region of interest (ROI) within each selected city by drawing a rectangle of approximately 50 × 50 km around the city center in Google Earth. Second, polygons enclosing different LCZ classes were manually delineated in Google Earth. These polygons are the preliminary labels. Afterward, Landsat 8 images covering the ROI were prepared.

After this preparation, an RF classifier was trained using the Landsat 8 images and the preliminary LCZ labels to produce an LCZ classification map of the specific city. This classification map and the satellite image on Google Earth served as auxiliary data for cross-checking the correctness and completeness of the LCZ labels. The details are explained as follows.

- **Correctness**: The crew visually inspected discrepancies between the classification maps and the labels of the polygons. If a mismatch was found for a labeled polygon, the crew inspected the satellite image on Google Earth and corrected the given label if necessary. This process was repeated until no noticeable discrepancy between the classification map and label was found.

- **Completeness**: The labeling crew cross-checked the classification result with the satellite image on Google Earth in unlabeled areas to find negative samples. For example, dense forest might be classified as water because of the lack of a dense forest label. The labeling crew then labeled those negative samples of dense forest and included them in the whole label data set. This hard-negative mining procedure was carried out iteratively until no noticeable discrepancies between the classification maps and Google Earth images in unlabeled areas were found.

The classification maps produced during the manual labeling process were employed only to provide guidance to the labeling crew and were not used in the final data. All LCZ labels in the final provided reference data fully relied on manual human annotation.

**FIGURE 4. The four-phase labeling project, showing the (a) learning, (b) labeling, (c) first validation, and (d) second validation phases.**
**Decision Rule of Local-Climate-Zone Labeling**

The decision rule consists of seven hierarchical questions:

A) Is it homogeneous for at least five pixels of 100 × 100 m?
B) Is the building footprint large?
C) Does any obvious industrial feature exist (such as oil tanks, cranes, or conveyor belts)?
D) What is the building height?
   - D1. Up to three floors
   - D2. Three to 10 floors
   - D3. Ten floors and higher.
E) For D1, what is the building surface fraction?
   - E1. Between 20% and 40%
   - E2. Smaller than 20%
   - E3. Between 40% and 70%
   - E4. Light material built with a surface fraction of larger than 60%
F) For D2, is the building surface fraction larger than 40%?
G) For D3, is the building surface fraction larger than 40%?

The percentage is estimated by experts with a 100 × 100-m polygon drawn on Google Earth. The building height is decided by experts using any available information, such as a 3D model, satellite images, or photos (Figure S1).

**FIGURE S1.** A flowchart of the labeling decision rule, which identifies one scene with seven decisions.

**VISUAL QUALITY-CONTROL PHASE**

Despite a clear quantitative definition that was agreed on by the labeling crew in the learning phase, personal bias and outliers still existed in the labeling result. Therefore, a manual inspection was required before quantitative validation to adjust personal biases and decrease the inevitable human mistakes. After the labeling phase, two persons other than the individual who labeled the polygons sequentially and independently validated the labels, as demonstrated in step C of Figure S1. These two persons were responsible for visually inspecting two types of signals in the classification map: 1) obvious outliers, such as water being classified as a dense high-rise building, and 2) a normal compactness-centric pattern of urban areas, that is, the compactness of urban buildings decreases from the city center toward the suburbs. If the obvious outliers cover a comparatively large area, a polygon with the correct label must be added. If an abnormal compactness pattern appears, the validation requires a detailed inspection, which often leads to adding polygons or correcting labels of existing polygons. We found that visual validation gave us a significant indication of label quality.

**LABEL POSTPROCESSING**

After obtaining labeled LCZ polygons, we discovered that the following postprocessing procedures were necessary.

- **Polygon shrinking:** Although all of the polygons were correctly labeled, some polygons in a given LCZ class were drawn in close proximity to another LCZ class. This might cause erroneous labels on the pixels close to the borders of the polygon when the polygon is rasterized, especially when using a large ground-sampling distance...
Class balancing: For those vector-format polygon labels in machine learning of EO images to be used, they must be rasterized into image format in certain geographic coordinate systems. We used GeoTIFF format and local Universal Transverse Mercator (UTM) coordinates. However, the polygons of nonurban LCZ classes (i.e., classes A–G) tend to be much larger in area than those of urban classes, because the percentage of nonurban areas is naturally larger and they are much easier for humans to label. This results in many more pixels (samples) for nonurban classes. To balance the number of samples among all of the LCZ classes, for each city, we reduced the number of samples of each of the nonurban classes (A–G) to \(N_m\), where \(N_m\) is the maximum number of samples from the urban classes (i.e., classes 1–10). If the number of samples of certain nonurban classes was less than \(N_m\), those classes remained untouched. The samples of the urban classes were not reduced because they are difficult to label. In this way, we were able to balance the different LCZ classes.

### Quantitative Quality Control and Validation Phase

The maximum accuracy achievable by any supervised learning procedure depends not only on the chosen algorithm but also on the quality of the training data. Therefore, we conducted quantitative evaluation on 10 European cities in the data set by having a group of remote sensing experts cast 10 independent votes on each labeled polygon to assess human labeling accuracy and identify possible remaining errors. Despite the huge labor cost, we believe this is essential for EO data and products to provide an error bar to users. This label evaluation procedure is discussed in detail in the "Label Evaluation" section.

### Preparing the Sentinel-1 Data

The Sentinel-1 mission provides an open access global SAR data set. We used the Sentinel-1 VV–VH dual-Pol single-look complex (SLC) level 1 data via the Copernicus Open Access Hub [50] employing an automatic script developed by the authors based on SentinelSat [51].

A series of preprocessing steps was applied to the Sentinel-1 data using the graph processing tool in the European Space Agency’s Sentinel Application Platform (SNAP) toolbox. The detailed configurations of the preprocessing are listed as follows.

- **Apply orbit profile**: This module downloads the latest released orbit profile so that a precisely geocoded product can be achieved.
- **Radiometric calibration**: Radiometry is employed to compute the backscatter intensity using sensor calibration parameters in the metadata. The output is set to a complex-valued image to preserve the relative phase between VV and VH channels.
- **TOPSAR deburst**: For each polarization channel, the Sentinel-1 1W product has three swaths. Each swath image consists of a series of bursts. The TOPSAR deburst merges all of these bursts and swaths into a single SLC image.
- **Polarimetric speckle reduction**: Speckle reduction was conducted by using the SNAP-integrated refined Lee filter. An unfiltered version is also included in the data set.
- **Terrain correction**: Terrain correction eliminates the distortion introduced by topographical variations. To accomplish the correction, the SRTM digital elevation model was used to provide height information. The data were resampled
to a 10-m GSD by nearest-neighbor interpolation. The data were geocoded into the WGS84/UTM coordinate system of the corresponding city with a GSD of 10 m.

To summarize, the Sentinel-1 data in the So2Sat LCZ42 data set contain the following eight real-valued bands:

- real part of the unfiltered VH channel
- imaginary part of the unfiltered VH channel
- real part of the unfiltered VV channel
- imaginary part of the unfiltered VV channel
- intensity of the refined Lee-filtered VH channel
- intensity of the refined Lee-filtered VV channel
- real part of the refined Lee-filtered covariance matrix off-diagonal element
- imaginary part of the refined Lee-filtered covariance matrix off-diagonal element.

**PREPARING THE SENTINEL-2 DATA**

We employed Google Earth Engine (GEE) to create cloud-free Sentinel-2 images [44]. The overall workflow, based on the GEE Python application programming interface, consisted of the following three main steps.

- **Querying step:** loading Sentinel-2 images from the catalogue
- **Scoring step:** calculating a cloud-related quality score for each loaded image
- **Mosaicking step:** mosaicking the selected images based on the meta-information generated in the preceding modules.

More details can be found in [45].

Sentinel-2 images contain bands B2, B3, B4, and B8 with 10-m GSD; bands B5, B6, B7, B8a, B11, and B12 with 20-m GSD; and bands B1, B9, and B10 with 60-m GSD. In the So2Sat LCZ42 data set, the 20-m bands were upsampled to 10-m GSD, and bands B1, B9, and B10 were discarded because they mostly contain data related to the atmosphere and, thus, bear little relevance to LCZ classification. To summarize, the Sentinel-2 data in the So2Sat LCZ42 data set contain the following 10 real-valued bands:

- band B2, 10-m GSD
- band B3, 10-m GSD
- band B4, 10-m GSD
- band B5, upsampled to 10 m from 20-m GSD
- band B6, upsampled to 10 m from 20-m GSD
- band B7, upsampled to 10 m from 20-m GSD
- band B8, 10-m GSD
- band B8a, upsampled to 10 m from 20-m GSD
- band B11, upsampled to 10 m from 20-m GSD
- band B12, upsampled to 10 m from 20-m GSD.

**CONTENT OF THE SO2SAT LCZ42 DATA SET**

By projecting the labels to the coregistered Sentinel-1 and Sentinel-2 images, we can extract Sentinel-1 and Sentinel-2 image patch pairs with the corresponding LCZ labels. We define the dimension of the image patches in the So2Sat LCZ42 data set as 32 × 32 pixels, which corresponds to a physical dimension of 320 × 320 m. To create nonoverlapping patches, we sampled the labeled polygons with a 320 × 320-m grid, where the grid nodes are the center of each image patch. We obtained 400,673 pairs of Sentinel image patches. The volume of the whole data set is approximately 56 GB.

For machine-learning purposes, the data set was split into a training set, a testing set, and a validation set, which consist of 352,366; 24,188; and 24,119 pairs of image patches, respectively. The training set comprises all of the image patches of 32 cities plus the 10 add-on areas in the city list. See “City List of the So2Sat LCZ42 Data Set” for the full list of cities). The remaining 10 cities are distributed across all of the continents and culture regions of the world. For each, we split the labels of each LCZ class into the western and eastern halves of a city to form the testing and validation sets, respectively. Therefore, all three data subsets are geographically separated from each other, despite our having drawn the testing and validation sets from the same list of cities.

**LABEL EVALUATION**

The maximum accuracy achievable by any supervised learning procedure depends not only on the chosen algorithm but also on the quality of the training data [46]. In the Human Influence Experiment (HUMINEX), Bechtel et al. [28] recently showed the difficulties associated with having human experts assign LCZ classes. Therefore, evaluating the labels that are the result of human expert knowledge is of vital importance for further use of the data set in the training of classification algorithms for large-scale automatic LCZ mapping.

**THE EVALUATION SET**

For the evaluation, we chose a subset of 10 European cities (Table 2) from the group of cities we labeled. The choice was based on the following three rationales:

- All of our labeling experts have lived in Europe for a significant number of years. This ensures familiarity with the general morphological appearance of European cities.
- Google Earth provides detailed 3D models for the 10 cities, which is of great help in determining the approximate heights of urban objects. This is necessary to be able to distinguish among the low-rise, mid-rise, and high-rise classes.

**City List of the So2Sat LCZ42 Data Set**

Cities used for training: Amsterdam, Beijing, Berlin, Bogota (added on), Buenos Aires (added on), Cairo, Cape Town, Caracas (added on), Changsha, Chicago (added on), Cologne, Dhaka (added on), Dongying, Hong Kong, Islamabad, Istanbul, Karachi (added on), Kyoto, Lima (added on), Lisbon, London, Los Angeles, Madrid, Manila (added on), Melbourne, Milan, Nanjing, New York, Paris, Philadelphia (added on), Qingdao, Rio de Janeiro, Rome, Salvador (added on), Sao Paulo, Shanghai, Shenzhen, Tokyo, Vancouver, Washington (D.C.), Wuhan, Zurich.

Cities used for testing and validation: Guangzhou, Jakarta, Moscow, Mumbai, Munich, Nairobi, San Francisco, Santiago de Chile, Sydney, Tehran.
As previously mentioned, LCZ labeling is very labor intensive. Reducing the evaluation set to 10 cities allowed us to generate more individual votes per polygon for better statistics. Unfortunately, few European cities contain LCZ class 7 (lightweight, low rise), which mostly describes informal settlements (e.g., slums). Therefore, we included the polygons of class 7 for an additional nine cities that are representative of the nine major non-European geographical regions of the world (Table 3).

**EVALUATION STRATEGY AND RESULTS**

For the evaluation experiment, 10 remote sensing experts (hereafter referred to as the label validation crew), who were already trained in applying the LCZ scheme to annotate urban areas, were provided with .km files containing the polygons of the original So2Sat LCZ42 data set, but without labels. They were then asked to reassess an LCZ class to every polygon, using Google Earth as the labeling environment. After all of the relabeled .km files were submitted, majority voting was carried out for each polygon; that is, each polygon was reassigned to the class for which a majority of the label validation crew had voted, although we kept the original label in case there was a draw between this original class and another major class. The polygon-wise and pixel-wise confusion matrices between these final labels and the votes of the label validation crew can be seen in Figure 5(c) and (d).

**INTERPRETATION OF THE EVALUATION RESULTS**

The confusion matrices in Figure 5 show the following.

- There is no significant difference between the polygon-wise and pixel-wise results, which indicates that the polygons are evenly distributed with respect to size.
- The majority voting step helped slightly improve the label confidences. Before the refinement, 11 of the 17 LCZ classes provided a confidence of more than 80%; after the refinement, this confidence level held for 13 classes.
- In general, confusion among the urban classes is slightly higher than among the nonurban classes.
- The classes with the most confidence are 8 (large, low rise), 7 (lightweight, low rise), and C (bush, scrub), with classes 2 (compact, mid rise) and E (bare rock/paved) following close behind.
- The classes with the least confidence are 3 (compact, low rise), 7 (lightweight, low rise), and C (bush, scrub), with classes 4 (open, high rise) and 9 (sparsely built) following behind. The main sources of confusion for these classes are summarized in Table 4.

These experiences go hand in hand with the findings of Bechtel et al. [28], who also found that LCZ classes A (dense trees), D (low plants), G (water), 2 (compact, mid rise), 6 (open, low rise), and 8 (large, low rise) were recognized consistently well by all operators, whereas classes 9 (sparsely built) and B (scattered trees) were reported as difficult to classify. Classes 1 (compact, high rise), 4 (open, high rise), 7 (lightweight, low-rise), and C (bush, scrub) were not present in most of their study cities and, thus, not discussed in detail.

Based on the major sources of confusion summarized in Table 4, all of these discrepancies appear fairly reasonable: apparently, it is difficult even for human experts to distinguish the vaguely defined characteristics “open” and “compact” as well as “mid rise” and “high rise.” In addition, sparsely built environments are understandably frequently confused with open low-rise neighborhoods, as is bush/scrubland with scattered trees and low plants.

Given the accordance with the findings of Bechtel et al. [28], the semantic subtleties of the LCZ classification scheme as well as a mean class confidence of approximately 80%

---

**TABLE 2. THE 10 EUROPEAN CITIES SELECTED FOR QUANTITATIVE LABEL EVALUATION.**

| CITY      | COUNTRY       |
|-----------|---------------|
| Amsterdam | The Netherlands |
| Berlin    | Germany       |
| Cologne   | Germany       |
| London    | United Kingdom |
| Madrid    | Spain         |
| Milan     | Italy         |
| Munich    | Germany       |
| Paris     | France        |
| Rome      | Italy         |
| Zurich    | Switzerland   |

**TABLE 3. THE ADDITIONAL NINE CITIES WHOSE POLYGONS OF CLASS 7 (LIGHTWEIGHT, LOW RISE) WERE USED FOR THE EVALUATION.**

| CITY                  | GEOGRAPHIC REGION      |
|-----------------------|------------------------|
| Guangzhou, China      | East Asia              |
| Islamabad, Pakistan   | Middle East            |
| Jakarta, Indonesia    | Southeast Asia         |
| Los Angeles, United States | North America   |
| Melbourne, Australia  | Oceania                |
| Moscow, Russia        | Eastern Europe         |
| Mumbai, India         | Indian subcontinent    |
| Nairobi, Kenya        | Sub-Saharan Africa     |
| Rio de Janeiro, Brazil| Latin America          |
before refinement by majority voting and 85% after refinement, the So2Sat LCZ42 data set can be considered a reliable source of labels for the training of machine-learning procedures aimed at automated LCZ mapping on a larger scale.

**BASELINE CLASSIFICATION ACCURACY**

To provide a baseline for achievable LCZ classification accuracy, we performed classification on the So2Sat LCZ42 data set using popular classifiers, including classical RFs, SVMs [30], and an attention-based ResNeXt, as proposed in [47] and [48]. The employed RF consists of 200 trees, and the max_depth is set to 10, with the other parameters set to the default. A radial basis function kernel is chosen for the SVM in the experiment. The depth of the ResNeXt is 29 layers, and the convolutional block attention module is plugged into each of the residual blocks. For RF and SVM, the pixel values of the patches are converted into vectors, using the statistical measures (maximum, minimum, standard deviations, and mean) of each band. All of the classifiers are trained using the training set and tested on the validation set.

The resulting accuracy based on the *Sentinel-2* images in the So2Sat LCZ42 data set can be seen in Table 5. The accuracy measures include overall accuracy, averaged accuracy, and the kappa coefficient. In addition, weighted accuracy, introduced in [28], is considered, because it gives user-defined weights to confusions between different classes. For example, misclassifying compact high rise as compact

| Original Label | Label Validation Votes | Final Label (After Majority Voting) |
|----------------|------------------------|-------------------------------------|
| 1              | 1234567890           | 1234567890                          |
| 2              | 1234567890           | 1234567890                          |
| 3              | 1234567890           | 1234567890                          |
| 4              | 1234567890           | 1234567890                          |
| 5              | 1234567890           | 1234567890                          |
| 6              | 1234567890           | 1234567890                          |
| 7              | 1234567890           | 1234567890                          |
| 8              | 1234567890           | 1234567890                          |
| 9              | 1234567890           | 1234567890                          |
| 10             | 1234567890           | 1234567890                          |

| Original Label | Label Validation Votes | Final Label (After Majority Voting) |
|----------------|------------------------|-------------------------------------|
| 1              | 1234567890           | 1234567890                          |
| 2              | 1234567890           | 1234567890                          |
| 3              | 1234567890           | 1234567890                          |
| 4              | 1234567890           | 1234567890                          |
| 5              | 1234567890           | 1234567890                          |
| 6              | 1234567890           | 1234567890                          |
| 7              | 1234567890           | 1234567890                          |
| 8              | 1234567890           | 1234567890                          |
| 9              | 1234567890           | 1234567890                          |
| 10             | 1234567890           | 1234567890                          |

| Original Label | Label Validation Votes | Final Label (After Majority Voting) |
|----------------|------------------------|-------------------------------------|
| 1              | 1234567890           | 1234567890                          |
| 2              | 1234567890           | 1234567890                          |
| 3              | 1234567890           | 1234567890                          |
| 4              | 1234567890           | 1234567890                          |
| 5              | 1234567890           | 1234567890                          |
| 6              | 1234567890           | 1234567890                          |
| 7              | 1234567890           | 1234567890                          |
| 8              | 1234567890           | 1234567890                          |
| 9              | 1234567890           | 1234567890                          |
| 10             | 1234567890           | 1234567890                          |

![FIGURE 5](image-url) The confusion matrices (values as percentages) of the original and final labels (refined by majority voting) versus the votes cast by the label validation crew for the polygons of the evaluation cities selected in Tables 2 and 3: (a) polygon-wise assessment of the original labels, (b) pixel-wise assessment of the original labels, (c) polygon-wise assessment of the final labels, and (d) pixel-wise assessment of final labels.
middle rise is less critical than mistaking compact high rise for water and, thus, should be penalized less.

**DISCUSSION**

The goal of this article is to provide documentation about a large benchmark data set for LCZ classification from *Sentinel-1* and *Sentinel-2* satellite data. Since the *Sentinel* data are openly available for the whole globe, the main intention of the data set is to enable the training of models that can be generalized to any unseen areas of the world. This is ensured by sampling the data from 52 cities located on all inhabited continents. In spite of these promising characteristics, two major challenges must be noted.

First, LCZs are sometimes hard to distinguish. As the label validation results shown in the "Label Evaluation section" illustrate, it is extremely difficult to distinguish some LCZ classes, even if human experts investigate several data sources (such as high-resolution optical imagery and 3D building models like those available in Google Earth). This especially holds for the distinction of different height levels in compact areas, but it is also true for open areas, which comprise both open land/vegetation and building structures. This must be acknowledged as a natural limitation when tackling LCZ mapping with remote sensing data. The limitation can be solved only by combining remote sensing data with other data sources, such as information from social media data.

Second, learning a generic LCZ prediction model is challenging. As described in the "Content of the So2Sat LCZ42 Data Set" section, the test and training sets are completely disjoint, with the test cities being distributed across the 10 major cultural regions of the inhabited world. Therefore, results achieved on this data set can be considered a good measure of how well the trained model would generalize to completely unseen data. In this regard, overall accuracies between 50% and 60% can already be considered promising—especially for a target scheme comprising 17 difficult-to-distinguish classes. Nevertheless, there is still room for improvement, as usually an accuracy of at least approximately 85% to 90% is required for land-cover mapping purposes, according to Anderson [49].

We hope that the community is eager to tackle these challenges and puts the So2Sat LCZ42 data set to good use to achieve significant progress in the global mapping of cities into LCZs.

**CONCLUSIONS AND OUTLOOK**

This article introduced a unique data set that contains manually labeled LCZ reference data as well as coregistered *Sentinel-1* and *Sentinel-2* image patch pairs for more than 42 cities and 10 smaller areas across the six inhabited continents on Earth. The article described the carefully designed labeling process and a rigorous evaluation procedure that ensures the quality of the data set. Despite the fact that each LCZ class is quantitatively defined in the original article, we discovered that several LCZ classes can be easily confused with each other, because the height and percentage of pervious surface of these classes cannot be easily distinguished by the human eye from aerial images during labeling. This renders the entire labeling process highly labor intensive.

Still, we were able to achieve an average class confidence of 85% through our human evaluation procedure with independent voting by 10 experts. Hence, this data set is a reliable source for the training of machine-learning procedures, and it can be considered a challenging and large-scale data fusion and classification benchmark data set for cutting-edge machine-learning methodological developments. Examples for possible research directions include the following.

- Since we have provided the label confusion matrix, the question of how to introduce such prior knowledge into machine learning, deep learning models in particular, is an interesting direction.
- Due to culture-induced diversity existing in the data, transferability of the models will be a key to achieving good classification results on a global scale.
- Radar and optical data have completely different yet partially complementary characteristics. Developing methods to fuse them in an optimal way or select appropriate features from such diverse data sources is of general interest to the remote sensing community.
- Thanks to the large scale of the proposed benchmark data set, it can serve as a test bed for the development of efficient training techniques.

Our vision in the near future is to produce a global LCZ classification map using multisensory remote sensing images, which will be made available to the community. Such geographic information seems trivial for developed
countries, but it is still very scarce on the global scale. For example, the city of Lagos, Nigeria (population: 21 million), does not have a quality 3D city model. Therefore, a quality LCZ classification map will become the firsthand information source for urban building volume and distribution. A global LCZ map will strongly boost urban geographic research and help us develop a better understanding of global urbanization. For this purpose, we invite everyone to contribute by using this data set and developing new, sophisticated algorithms.

ACKNOWLEDGMENTS
This research was funded by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant ERC-2016-StG-714087 (So2Sat, project website: http://www.so2sat.eu/), Helmholtz Association under the framework of the Helmholtz Artificial Intelligence Cooperation Unit–Local Unit “Munich Unit @Aeronautics, Space and Transport (MAS-Tri),” and Helmholtz Excellent Professorship “Data Science in Earth Observation–Big Data Fusion for Urban Research.”

AUTHOR INFORMATION
Xiao Xiang Zhu (xiaoxiang.zhu@dlr.de) received her M.Sc. degree, Dr.-Ing. degree, and habilitation in the field of signal processing from the Technical University of Munich (TUM), Germany, in 2008, 2011, and 2013, respectively. She is currently a professor for signal processing in Earth observation (EO) at TUM and the head of the department’s EO Data Science at the Remote Sensing Technology Institute, German Aerospace Center. Since 2019, she has been co-coordinating the Munich Data Science Research School. She is also leading the Helmholtz Artificial Intelligence Cooperation Unit–Research Field “Aeronautics, Space and Transport.” Her main research interests are remote sensing and EO, signal processing, and machine learning and data science, with a special application focus on global urban mapping. She is an associate editor of IEEE Transactions on Geoscience and Remote Sensing. She is a Senior Member of the IEEE.

Jingliang Hu (jingliang.hu@dlr.de) received his B.Sc. degree in geographic information systems from the Chengdu University of Technology, China, in 2011, the M.Sc. degree in cartography and geographic information system from Wuhan University, China, in 2014, and the M.Sc. degree in space science and technology from the Technical University of Munich (TUM), Germany, in 2014, where he is currently pursuing the Ph.D. degree in signal processing in Earth observation. Since July 2015, he has been a research associate with the Remote Sensing Technology Institute, German Aerospace Center, Germany. His research interests include topological data analysis, data fusion, machine learning, and deep learning.

Chunping Qiu (chunping.qiu@tum.de) is with Signal Processing in Earth Observation, Technical University of Munich, Germany.

Yilei Shi (yilei.shi@tum.de) is with the chair of Remote Sensing Technology, Technical University of Munich, Germany.

Jian Kang (jian.kang@tum.de) was with Signal Processing in Earth Observation, Technical University of Munich, Germany when conducting the work reported in this article.

Lichao Mou (lichao.mou@dlr.de) is with Signal Processing in Earth Observation, Technical University of Munich, Germany.

Hossein Bagheri (hossein.bagheri@tum.de) is with Signal Processing in Earth Observation, Technical University of Munich, Germany.

Matthias Häberle (matthias.haeberle@dlr.de) is with the Remote Sensing Technology Institute, German Aerospace Center.

Yuansheng Hua (yuansheng.hua@tum.de) is with the Remote Sensing Technology Institute, German Aerospace Center.

Rong Huang (rong.huang@tum.de) was with Signal Processing in Earth Observation, Technical University of Munich, Germany, when conducting the work reported in this article.

Lloyd Hughes (lloyd.hughes@tum.de) is with Signal Processing in Earth Observation, Technical University of Munich, Germany.

Hao Li (hao.li@uni-heidelberg.de) was with Signal Processing in Earth Observation, Technical University of Munich, Germany, when conducting the work reported in this article.

Yao Sun (yao.sun@dlr.de) is with the Remote Sensing Technology Institute, German Aerospace Center.

Guichen Zhang (guichen.zhang@tum.de) was with Signal Processing in Earth Observation, Technical University of Munich, Germany, when conducting the work reported in this article.

Shiyao Han (shiyao.han@tum.de) was with Signal Processing in Earth Observation, Technical University of Munich, Germany, when conducting the work reported in this article.

Michael Schmitt (m.schmitt@tum.de) received his Dipl.-Ing. (university) degree in geodesy and geoinformation, his Dr.-Ing. degree in remote sensing, and his habilitation in data fusion from the Technical University of Munich (TUM), Germany, in 2009, 2014, and 2018, respectively. Since 2015, he has been a senior researcher and deputy head at the professorship for Signal Processing in Earth Observation at TUM; in 2019, he was additionally appointed as an adjunct teaching professor at the Department of Aerospace and Geodesy of TUM. In 2016, he was a guest scientist at the University of Massachusetts, Amherst. His research focuses on image analysis and machine learning applied to the extraction of information from remote-sensing observations. In particular, he is interested in remote sensing data fusion with a focus on synthetic aperture radar and optical data. He is a cochair of the International Society for Photogrammetry and Remote Sensing Working Group 1/3 “SAR and Microwave Sensing.” He frequently serves as a reviewer for a number of renowned international journals and conferences and has received several Best Reviewer awards. He...
is an associate editor of *IEEE Geoscience and Remote Sensing Letters*. He is a Senior Member of the IEEE.

**Yuanyuan Wang** (ywang@tum.de) received his B.Eng. degree (with honors) in electrical engineering from Hong Kong Polytechnic University in 2008 and his M.Sc. and Dr. Ing. degrees from the Technical University of Munich (TUM), Germany, in 2010 and 2015, respectively. In June and July 2014, he was a guest scientist with the Institute of Visual Computing, ETH Zürich, Switzerland. He is currently with the Signal Processing in Earth Observation (EO), TUM, and the Department of EO Data Science, German Aerospace Center, Wessling, Germany, where he leads the working group Big SAR Data. His research interests include optimal and robust parameters estimation in multibaseline interferometric synthetic aperture radar (SAR) techniques, multisensor fusion algorithms of SAR and optical data, nonlinear optimization with complex numbers, machine learning in SAR, and high-performance computing for big data. He received a Best Reviewer award for *IEEE Transactions on Geoscience and Remote Sensing* in 2016. He is a Member of the IEEE.

**REFERENCES**

[1] H. Taubenböck, T. Esch, A. Felbier, M. Wiesner, A. Roth, and S. Dech, “Monitoring urbanization in megacities from space,” *Remote Sens. Environ.*, vol. 117, pp. 162–176, 2012. doi: 10.1016/j.rse.2011.09.015.

[2] “World urbanization prospects: 2018 revision,” United Nations, New York, 2018.

[3] T. Esch et al., “TanDEM-X mission—new perspectives for the inventory and monitoring of global settlement patterns,” *J. Appl. Remote Sens.*, vol. 6, no. 1, p. 061702, 2012. doi: 10.1117/1.JRS.6.061702.

[4] T. Esch et al., “Urban footprint processor: Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1617–1621, 2013. doi: 10.1109/LGRS.2013.2272953.

[5] M. Pesaresi et al., “A global human settlement layer from optical HR/VHR RS data: Concept and first results,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2102–2131, 2013. doi: 10.1109/JSTARS.2013.2271445.

[6] Y. Ban, P. Gong, and C. Gini, “Global land cover mapping using earth observation satellite data: Recent progresses and challenges,” *ISPRS J. Photogramm. Remote Sens.*, vol. 103, no. 1, pp. 1–6, 2015. doi: 10.1016/j.isprsjprs.2015.01.001.

[7] I. D. Stewart, “Local climate zones: Origins, development, and application to urban heat island studies,” presented at the Annu. Meeting American Association of Geographers, Seattle, WA, Apr. 11–16, 2011.

[8] B. Bechtel et al., “Mapping local climate zones for a worldwide database of the form and function of cities,” *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 1, pp. 199–219, 2015. doi: 10.3390/ijgi4010199.

[9] I. D. Stewart and T. R. Oke, “Local climate zones for urban temperature studies,” *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 12, pp. 1879–1900, 2012. doi: 10.1175/BAMS-D-11-00019.1.

[10] I. D. Stewart, T. R. Oke, and E. S. Krainenhoff, “Evaluation of the ‘local climate zone’ scheme using temperature observations and model simulations,” *Int. J. Climatol.*, vol. 34, no. 4, pp. 1062–1080, 2014. doi: 10.1002/joc.3746.

[11] D. Fenner, F. Meier, B. Bechtel, M. Otto, and D. Scherer, “Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany,” *Meteorol. Z.*, vol. 26, no. 5, pp. 525–547, 2017. doi: 10.1127/metez/2017/0861.

[12] S. J. Quan, F. Dutt, E. Woodworth, Y. Yamagata, and P. P.-J. Yang, “Local climate zone mapping for energy resilience: A fine-grained and 3D approach,” *Energy Procedia*, vol. 105, pp. 3777–3783, May 2017. doi: 10.1016/j.egypro.2017.03.883.

[13] J. A. Quanz, S. Ulich, D. Fenner, A. Holtmann, and J. Eismann, “Micro-scale variability of air temperature within a local climate zone in Berlin, Germany, during summer,” *Climate*, vol. 6, no. 1, p. 5, 2018. doi: 10.3390/cli6010005.

[14] R. Koitharkar and A. Bagade, “Evaluating urban heat island in the critical local climate zones of an Indian city,” *Landsc. Urban Plan.*, vol. 169, no. 2018, pp. 92–104, 2018. doi: 10.1016/j.landurbplan.2017.08.009.

[15] J. Geletič, M. Lehner, S. Šavić, and D. Milošević, “Inter/intra-zonal seasonal variability of the surface urban heat island based on local climate zones in three central European cities,” *Build. Environ.*, vol. 156, pp. 21–32, June 2019. doi: 10.1016/j.buildenv.2019.04.011.

[16] C. B. Koc, P. Osmond, A. Peters, and M. Irger, “Understanding land surface temperature differences of local climate zones based on airborne remote sensing data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2724–2730, 2018. doi: 10.1109/JSTARS.2018.2815004.

[17] C. B. Koc, P. Osmond, A. Peters, and M. Irger, “Mapping local climate zones for urban morphology classification based on airborne remote sensing data,” in *Proc. 2017 IEEE Joint Urban Remote Sensing Event (JURSE)*, pp. 1–4. doi: 10.1109/JURSE.2017.7924611.

[18] J. Geletič, M. Lehner, P. Dobrovolný, and M. Žuvela-Aloise, “Spatial modelling of summer climate indices based on local climate zones: Expected changes in the future climate of Brno, Czech republic,” *Clim. Chang.*, vol. 152, nos. 3–4, pp. 487–502, 2019. doi: 10.1007/s10584-018-2353-5.

[19] A. Wicki and E. Parlow, “Attribution of local climate zones using a multitemporal land use/land cover classification scheme,” *J. Appl. Remote Sens.*, vol. 11, no. 2, p. 026001, 2017. doi: 10.1117/1.JRS.11.026001.

[20] H. C. Ho et al., “Spatial variability of geriatric depression risk in a high-density city: A data-driven socio-environmental vulnerability mapping approach,” *Int. J. Environ. Res. Public Health*, vol. 14, no. 9, p. 994, 2017. doi: 10.3390/ijerph14090994.

[21] N. Yokoya et al., “Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, 2018. doi: 10.1109/JSTARS.2018.2799698.

[22] WUDAPT. Accessed on: Nov. 16, 2018. [Online]. Available: http://www.wudapt.org/locz/locz-framework/

[23] G. Mills, J. Ching, L. See, B. Bechtel, and M. Foley, “An introduction to the WUDAPT project,” in *Proc. 9th Int. Conf. Urban Climate*, Toulouse, France, 2015, pp. 20–24.
[24] B. Bechtel et al., "CENSUS of cities: LCZ classification of cities (Level 0)—Workflow and initial results from various cities," in Proc. 9th Int. Conf. Urban Climate, Toulouse, France, 2015.

[25] J. Hidalgo et al., "Comparison between local climate zones maps derived from administrative datasets and satellite observations," Urban Clim., vol. 27, pp. 64–89, Mar. 2019. doi: 10.1016/j.uclim.2018.10.004.

[26] B. Bechtel et al., "Generating WUDAPT level 0 data—current status of production and evaluation," Urban Clim., vol. 27, pp. 24–45, Mar. 2019. doi: 10.1016/j.uclim.2018.10.001.

[27] M. Demuzere, B. Bechtel, A. Middel, and G. Mills, "Mapping Europe into local climate zones," PLoS ONE, vol. 14, no. 4, p. e0214474, 2019. doi: 10.1371/journal.pone.0214474.

[28] B. Bechtel et al., "Quality of crowdsourced data on urban morphology: The human influence experiment (HUMINEX)," Urban Sci., vol. 1, no. 2, p. 15, 2017. doi: 10.3390/urbansci1020015.

[29] C. Ren, R. Wang, M. Cai, Y. Xu, Y. Zheng, and E. Ng, "The accuracy of LCZ maps generated by the world urban database and access portal tools (WUDAPT) method: A case study of Hong Kong," in Proc. 4th Int. Conf. Countermeasure Urban Heat Islands, Singapore, 2016, pp. 1–11.

[30] Y. Xu, C. Ren, M. Cai, N. Y. Y. Edward, and T. Wu, "Classification of local climate zones using ASTER and Landsat data for high-density cities," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 7, pp. 3397–3405, 2017. doi: 10.1109/JSTARS.2017.2683484.

[31] C. Qiu, M. Schmitt, P. Ghamisi, and X. X. Zhu, "Effect of the training set configuration on Sentinel-2-based urban local climate zone classification," in Proc. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018, pp. 931–936. doi: 10.5194/isprs-archives-XLIII-2-931-2018.

[32] J. Hu, P. Ghamisi, and X. Zhu, "Feature extraction and selection of Sentinel-1 dual-pol data for global-scale local climate zone classification," ISPRS Int. J. Geo-Inf., vol. 7, no. 9, p. 379, 2018. doi: 10.3390/ijgi7090379.

[33] S. Sukhanov, I. Tankoyeu, J. Louradour, R. Heremans, D. Trofimova, and C. Debes, "Multilevel ensembling for local climate zones classification," in Proc. 2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), pp. 1201–1204. doi: 10.1109/IGARSS.2017.8127173.

[34] N. Yokoya, P. Ghamisi, and J. Xia, "Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning," in Proc. 2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), pp. 1197–1200. doi: 10.1109/IGARSS.2017.8127172.

[35] C. Xu, X. Zhu, N. Tapper, and B. Bechtel, "Urban climate zone classification using convolutional neural network and ground-level images," Prog. Phys. Geogr., Earth Environ., vol. 43, no. 3, pp. 410–424, 2019. doi: 10.1177/0309133319837711.

[36] 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, 2017. doi: 10.1109/MGRS.2017.2762307.

[37] C. Qiu, M. Schmitt, L. Mou, and X. X. Zhu, "Urban local climate zone classification with a residual convolutional neural network and multi-seasonal Sentinel-2 images," in Proc. 10th IAPR Workshop on Pattern Recognition in Remote Sensing, 2018, pp. 1–5. doi: 10.1109/PRRS.2018.8486155.

[38] C. Qiu, M. Schmitt, P. Ghamisi, L. Mou, and X. X. Zhu, "Feature importance analysis of Sentinel-2 imagery for large-scale urban local climate zone classification," in Proc. 2018 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS), pp. 4681–4684. doi: 10.1109/IGARSS.2018.8517732.

[39] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," ISPRS J. Photogramm. Remote Sens., vol. 154, pp. 151–162, Aug. 2019. doi: 10.1016/j.isprsjprs.2019.05.004.

[40] C. Yoo, D. Han, J. Im, and B. Bechtel, "Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images," ISPRS J. Photogramm. Remote Sens., vol. 157, pp. 155–170, Nov. 2019. doi: 10.1016/j.isprsjprs.2019.09.009.

[41] Y. Yu et al., "Mapping impervious surfaces in town–rural transition belts using Chinas GF-2 imagery and object-based deep CNNs," Remote Sens., vol. 11, no. 3, p. 280, 2019. doi: 10.3390/rs11030280.

[42] H. Jing et al., "Effective classification of local climate zones based on multi-source remote sensing data," in Proc. IGARSS 2019–2019 IEEE Int. Geoscience and Remote Sensing Symp., pp. 2666–2669. doi: 10.1109/IGARSS.2019.8898475.

[43] M. Demuzere, B. Bechtel, and G. Mills, "Global transferability of local climate zone models," Urban Clim., vol. 27, pp. 46–63, Mar. 2019. doi: 10.1016/j.uclim.2018.11.001.

[44] 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, July 2017. doi: 10.1016/j.rse.2017.06.031.

[45] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Aggregating cloud-free Sentinel-2 images with google earth engine," ISPRS Ann. Photogramm., Remote Sens. Spat. Inf. Sci., vol. IV-2/W7, pp. 145–152, Sept. 2019. doi: 10.5194/isprs-annals-IV-2-W7-145-2019.

[46] C. E. Brodley and M. A. Friedl, "Identifying mislabeled training data," J. Artif. Intell. Res., vol. 11, no. 1, pp. 131–167, 1999. doi: 10.1613/jair.606.

[47] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017, pp. 1492–1500. doi: 10.1109/CVPR.2017.634.

[48] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," in Proc. European Conf. Computer Vision, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.

[49] J. R. Anderson, "Land-use classification schemes: Used in selected recent geographic applications of remote sensing," Photogramm. Eng., vol. 37, no. 4, pp. 379–387, 1971.

[50] European Space Agency, "Copernicus open access hub." Accessed on: 2018. [Online]. Available: https://scihub.copernicus.eu/

sentinelst,sentinelst,GitHub. Accessed on: 2018. [Online]. Available: https://github.com/sentinelsat/sentinelst