remote sensing

MDPI

Article

Feature Importance Analysis for Local Climate Zone Classification Using a Residual Convolutional Neural Network with Multi-Source Datasets

Chunping Qiu [1], Michael Schmitt [1], Lichao Mou [1], Pedram Ghamisi [2] and Xiao Xiang Zhu [1,3,*]

[1] Signal Processing in Earth Observation, Technical University of Munich (TUM), 80333 Munich, Germany; chunping.qiu@tum.de (C.Q.); m.schmitt@tum.de (M.S.); lichao.mou@dlr.de (L.M.)
[2] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology (HIF), Exploration, D-09599 Freiberg, Germany; p.ghamisi@gmail.com
[3] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany
* Correspondence: xiaoxiang.zhu@dlr.de; Tel.: +49-(0)8153-28-3531

Received: 2 August 2018; Accepted: 27 September 2018; Published: 1 October 2018

Abstract: Global Local Climate Zone (LCZ) maps, indicating urban structures and land use, are crucial for Urban Heat Island (UHI) studies and also as starting points to better understand the spatio-temporal dynamics of cities worldwide. However, reliable LCZ maps are not available on a global scale, hindering scientific progress across a range of disciplines that study the functionality of sustainable cities. As a first step towards large-scale LCZ mapping, this paper tries to provide guidance about data/feature choice. To this end, we evaluate the spectral reflectance and spectral indices of the globally available Sentinel-2 and Landsat-8 imagery, as well as the Global Urban Footprint (GUF) dataset, the OpenStreetMap layers *buildings* and *land use* and the Visible Infrared Imager Radiometer Suite (VIIRS)-based Nighttime Light (NTL) data, regarding their relevance for discriminating different Local Climate Zones (LCZs). Using a Residual convolutional neural Network (ResNet), a systematic analysis of feature importance is performed with a manually-labeled dataset containing nine cities located in Europe. Based on the investigation of the data and feature choice, we propose a framework to fully exploit the available datasets. The results show that GUF, OSM and NTL can contribute to the classification accuracy of some LCZs with relatively few samples, and it is suggested that Landsat-8 and Sentinel-2 spectral reflectances should be jointly used, for example in a majority voting manner, as proven by the improvement from the proposed framework, for large-scale LCZ mapping.

Keywords: Local Climate Zones (LCZs); Sentinel-2; Landsat-8; spectral reflectance; classification; Residual convolutional neural Network (ResNet)

1. Introduction

Local Climate Zones (LCZs) have been established as an interdisciplinary scheme to describe urban morphology on a neighborhood scale [1]. The 17 LCZ classes are based on climate-relevant surface properties on the local-scale, mainly related to 3D surface structure (e.g., height and density of buildings and trees), surface cover (e.g., vegetation or paved), as well as anthropogenic (anthropogenic heat output) parameters. The scheme contains ten "built" and seven "natural" classes, which are depicted in Figure 1, intended to be universal and applicable in cities all over the world, offering the possibility to compare different areas of different cities with trenchant distinctions representing the heterogeneous thermal behavior within an urban environment [2].

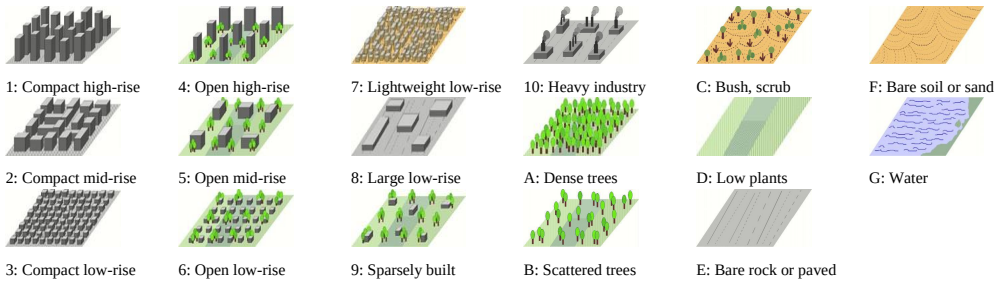


Figure 1. Visualization of the LCZ concept [1].

Besides the increasing impact on urban climate science worldwide [3–8], recently, researchers have started to use the LCZ scheme to classify the internal structure of urban areas, providing promising information for various applications such as infrastructure planning, disaster mitigation and population assessment [9] in this increasingly urbanized world [10]. Recently, it has been shown that useful insights can also be gained by investigating land surface temperature in different LCZs [11], even though LCZs are associated with air temperature when being proposed. Besides, from reliable LCZ maps, detailed information on human settlements can be further extracted, which can directly contribute to monitoring, assessing and decision making regarding the 2030 Agenda for Sustainable Development and provide relevant data for the Sustainable Development Goals (SDG), specifically SDG 11 (Sustainable Cities and Communities, "Make cities and human settlements inclusive, safe, resilient, and sustainable.") [12]. In 2017, ref. [13] used the LCZ framework for monitoring sustainable urbanization and to assess the availability of adequate and safe housing, with a case study in the cities of Johannesburg and Pretoria in South Africa. Last but not least, as an environmental factor, LCZs are expected to enable evidence-based strategies for planning healthy and green cities worldwide [14].

Supervised classification with remote sensing data provides a valuable tool for automatic LCZ mapping, as illustrated by the existing literature [15]. However, global LCZ mapping is still challenging due to the limited number of high quality ground truths, as well as a large intra-class variability of spectral signatures caused by the regional variations of vegetation and artificial materials, as well as other variations in cultural and physical environmental factors [2].

For automatic large-scale LCZ classification, the challenges are data availability and high generalization ability, as well as the transferability of the employed classification algorithm. Available datasets with high potential for this task include, but are not limited to, imagery (e.g., Landsat-8, Sentinel-1 and Sentinel-2), vector data (OpenStreetMap (OSM)) [16], settlement layers (Global Urban Footprint (GUF)) [17–19] and VIIRS nighttime light [20–23]. It is of great importance to understand the specific potential of each of these datasets and features, which is a common topic in hyperspectral image analysis [24,25]. However, only little literature exists in this regard. The work in [26] investigated the feature importance for LCZ mapping, showing that NDVI is the most important feature among the spectral reflectance of Landsat-8, spectral indices extracted from the Landsat-8 channels and the OSM layers (*land use, building and water*).

As a first step for large-scale or even global LCZ mapping, we focus on the globally available imagery provided by the Sentinel-2 and Landsat-8 mission [27], as well as the GUF, OSM and VIIRS nighttime light layers, using a Residual convolutional neural Network (ResNet) [28,29], as a framework for our investigations. Our work intends to provide answers to the following questions:

- Which dataset is better suited for LCZ classification, Sentinel-2 or Landsat-8? How do the external auxiliary datasets (GUF, OSM layers and NTL) contribute to the LCZ classification?
- How does one choose a proper dataset and suitable input features for LCZ classification, and what is the achievable accuracy?
- What are the main challenges for LCZ classification, and what are the possible solutions?

The remainder of this paper is structured as follows: Section 2 describes our proposed experimental setup of feature importance analysis for LCZ classification, as well as the resulting analysis and discussions.

Based on the findings in Section 3, Section 4 proposes a framework to fully exploit the datasets and shows the comparative LCZ classification accuracy, as well as the produced LCZ maps. Finally, Section 5 answers the mentioned questions in Section 1 based on the achieved results, before Section 6 summarizes and concludes the work.

2. Feature Importance Analysis for LCZ Classification with Multi-Source Datasets

2.1. Study Areas and LCZ Dataset

Our study areas are spread over nine cities located in the heart of Europe, as depicted in Figure 2.

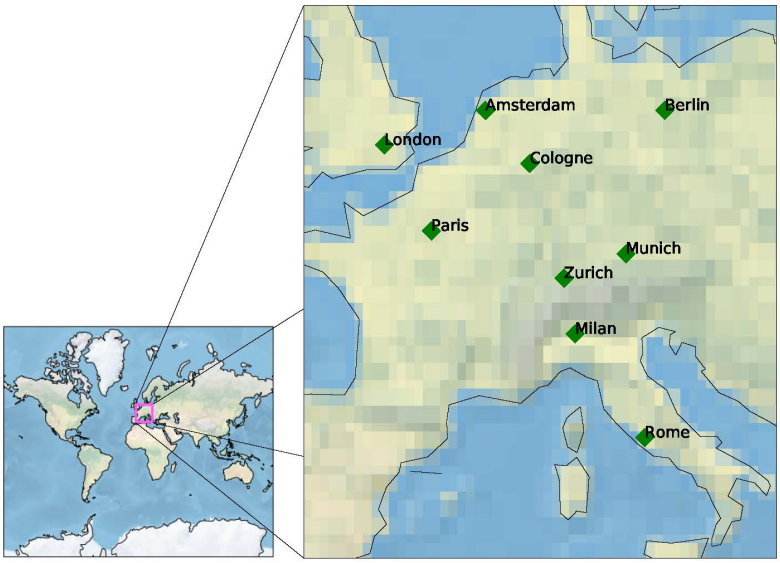


Figure 2. The nine test cities distributed across Europe.

The LCZ ground truth labels available for selected neighborhoods in the nine cities are taken from the LCZ42 dataset [30]. Figure 3 illustrates the variability of both the sample number and the class distribution among different cities. It should be noted that in these nine cities, LCZ Class 7 (lightweight low-rise), which mostly indicates slums, does not exist.

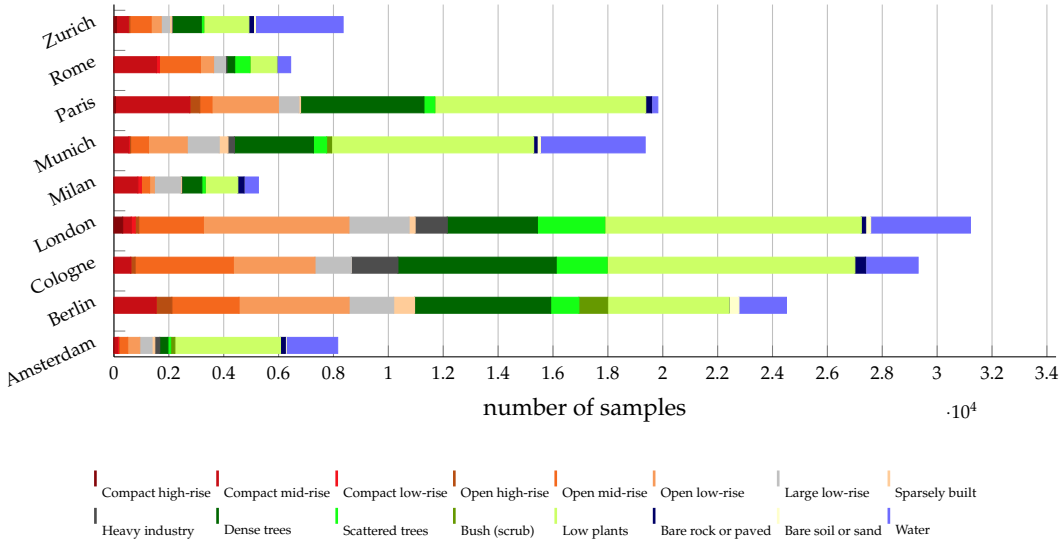


Figure 3. The sample number of the LCZ ground truth in the nine cities.

2.2. ResNet for LCZ Classification

For our investigations, we use a ResNet as the classifier, since ResNet has been shown to have superior classification performance [28]. By explicitly reformulating the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions, it has been shown that these residual networks are easier to optimize and can gain very competitive accuracy from considerably increased depth on the ImageNet dataset, compared to networks such as VGG and GoogleNet. The exact architecture of the fairly simple ResNet we train is shown in Figure 4. Overall, it has four residual blocks, and each of them consists of three convolutional layers and a shortcut connection that by-passes two stacked convolutional layers by performing identity mapping, which are then added together with the output of stacked convolutions. We utilize convolutional layers with a very small receptive field of 3×3, and the number of feature maps increases towards deeper blocks, doubling after each block. Max-pooling is performed over 2×2 pixel windows with a stride of two. For training the network, we use the TensorFlowframework. We choose Nesterov Adam as the optimization algorithm for our task, as it shows faster convergence than standard stochastic gradient descent with momentum. We fix the parameters of Nesterov Adam as recommended: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$ and a schedule decay of 0.004. We use a fairly small learning rate of 0.0002.

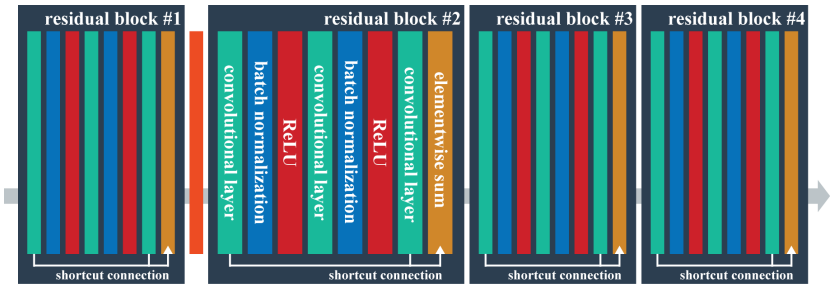


Figure 4. The architecture of the ResNet used for LCZ classification.

2.3. Input Datasets and Features

The input features being investigated in this paper are:

- Spectral reflectance:

  For each city, we have downloaded one cloud-free Sentinel-2 image and one Landsat-8 image from Google Earth Engine (GEE) [31]: Landsat-8 surface reflectance and Sentinel-2 MSI (TOA reflectance). Ten multispectral bands of Sentinel-2 imagery are used in this study: B2, B3, B4 and B8 with 10-m Ground Sampling Distance (GSD) and B5, B6, B7, B8a, B11 and B12 with 20-m GSD. The 20-m bands are up-sampled to 10-m GSD. The bands B1, B9 and B10 are not considered in this study because they contain mostly information about the atmosphere and thus bear little relevance to LCZ classification. Besides, nine multispectral bands of Landsat-8 imagery are also used: five Visible and Near-Infrared (VNIR) bands and two Short-Wave Infrared (SWIR) bands processed to orthorectified surface reflectance and two Thermal Infrared (TIR) bands processed to orthorectified brightness temperature. All Landsat-8 bands are up-sampled to 10-m GSD, in order to be aligned with Sentinel-2 images.

- Spectral indices:

  Spectral indices are extracted from both Sentinel-2 and Landsat-8 images. The well-established indices Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Modified Normalized Difference Water Index (MNDWI) [32], Normalized Difference Built Index (NDBI) [33], Normalized Built-up Area Index (NBAI), Band Ratio for Built-up Area (BRBA) and Bare-Soil Index (BSI) are also considered [34], since they can provide indications about vegetation, water, buildings, soil, etc. [2].

- Other auxiliary data:

  In addition, we were allowed to access DLR's Global Urban Footprint (GUF), a binary layer derived from TanDEM-X data, which indicates urban areas [19] globally. Besides, the Visible Infrared Imager Radiometer Suite (VIIRS)-based Nighttime Light (NTL) data are downloaded from GEE. Finally, we have downloaded the OpenStreetMap layers *buildings* and *land use* from OpenStreetMap Data Extracts (http://download.geofabrik.de/) for each city. As auxiliary data, GUF, NTL and OSM are re-sampled to 10-m GSD.

2.4. Setup of Feature Importance Analysis for LCZ Classification

In this section, a feature importance investigation has been designed with the purpose of: (i) evaluating the importance of the spectral reflectance and index measures from Sentinel-2 and Landsat-8 imagery with respect to the LCZ classification; (ii) assessing and comparing the applicability of the external auxiliary GUF and OSM layers to LCZ classification.

Table 1 shows the feature combinations being investigated in this study. Comparing the results from the configurations *S_1–S_4*, the relative importance of index measures, GUF, OSM and nighttime light can be interpreted, and the same holds for configurations *L_1–L_4*. Besides, a comparative analysis of all classification results can provide a

For all experiments, we rely on cross-validation, i.e., each time, samples from eight cities are used for training, while those from the remaining city are used for testing.

Table 1. Comparative experiment ID and the corresponding employed data and features. GUF, Global Urban Footprint; NTL, Nighttime Light.

Data and Feature	Dataset	
	Sentinel-2	**Landsat-8**
Spectral reflectance	*S_0*	*L_0*
Spectral reflectance, Indices	*S_1*	*L_1*
Spectral reflectance, GUF	*S_2*	*L_2*
Spectral reflectance, OSM	*S_3*	*L_3*
Spectral reflectance, NTL	*S_4*	*L_4*
Spectral reflectance, GUF, OSM	*S_5*	*L_5*

Besides the commonly-used classification measures Overall Accuracy (OA), Averaged Accuracy (AA), the kappa coefficient, etc., and another measure, Weighted Accuracy (WA) is also used for assessment. WA was introduced in [35], giving different weights to different misclassifications based on a systematical analysis of the potential climate impact of those misclassifications, taking into account the properties such as openness, height, cover and thermal inertia. For example, the misclassification between compact high rise and compact middle rise is less severe than that between compact high rise and water, and is thus penalized less.

3. Results of Feature Importance Analysis

Results from the framework described in Section 2 are shown and discussed in this section. Classification results of different data and feature configurations (as described in Table 1) are compared in Table 2.

By comparing the results from different configurations in Table 2, we can see the different contributions of index measures, GUF, OSM and NTL, respectively. This contribution difference for ResNet is $OSM > NTL > GUF > indices$. Besides, the combination spectral reflectance-OSM is the one configuration that provides the best accuracy for Landsat-8 imagery, and it also provides almost the best accuracy for Sentinel-2 imagery. Index measures contribute negatively to LCZ classification in this case, for both Sentinel-2 and Landsat-8 images. This may result from the feature extraction ability from the raw data of the employed ResNet, thus making the extracted index measures unnecessary.

In addition, Comparing the accuracy of *S_1*, *S_2* and *S_3* to *S_0*, it can be seen that no benefits can be achieved by GUF. This explains why the accuracy of *S_5* is only slightly better than *S_3*. This holds also true for Landsat-8 imagery. NTL provides better results than GUF, but it still does not improve the baseline accuracy where only the spectral reflectance is used, for both Sentinel-2 and Landsat-8. One possible reason is the course resolution of the NTL data (724 m in both directions). In addition, the correspondence between the value in NTL and the LCZs is not clear, so the contribution of NTL to the LCZs differentiation might be limited, especially considering that the experiments are designed in a cross-validation manner, where the training and test cities are completely different.

Table 2. Classification accuracy of different feature configurations, as explained in Table 1. The results are averaged over all 9 test cities. The values highlighted in bold are the highest ones for Sentinel-2 or Landsat-8. WA, Weighted Accuracy; AA, Averaged Accuracy.

Input	Sentinel-2						Landsat-8						Stacking
	S_0	***S_1***	***S_2***	***S_3***	***S_4***	***S_5***	***L_0***	***L_1***	***L_2***	***L_3***	***L_4***	***L_5***	***S_0 + L_0***
OA	**0.71**	0.63	0.67	**0.71**	0.68	**0.71**	0.72	0.58	0.70	**0.73**	0.71	**0.73**	0.72
WA	0.93	0.90	0.93	**0.94**	0.92	**0.94**	0.93	0.87	**0.94**	**0.94**	0.93	**0.94**	0.93
AA	0.46	0.41	0.45	0.49	0.46	**0.50**	0.48	0.37	0.47	**0.50**	0.48	**0.50**	0.48
Kappa	**0.65**	0.56	0.59	0.64	0.61	**0.65**	0.67	0.50	0.64	**0.68**	0.65	0.67	0.65

Table 2 shows that OSM can improve the performance of the classifier while GUF and NTL do not contribute much to the overall classification performance, regarding the generalization ability since the experiments are carried out in a cross-validation manner. However, how do these external auxiliary datasets contribute to the LCZ classification in detail? This can be further explained with Figure 5, where the feature importance for each LCZ has been shown. From Figure 5, we can see that, for Sentinel-2 images, OSM mainly contributes to the accuracy of large low-rise, heavy industry, dense trees, scattered trees and bush (scrub); GUF mainly contributes to the accuracy of compact high-rise and open high-rise, while NTL mainly contributes to the accuracy of compact low-rise, open high-rise, dense trees, scattered trees and bare rock or paved. For Landsat-8, OSM mainly contributes to compact high-rise, compact mid-rise, compact low-rise, heavy industry and bare soil or sand; GUF mainly contributes to open high-rise, sparsely-built, bush (scrub), bare rock or paved and bare soil or sand, while NTL mainly contributes to the accuracy of open high-rise, bush (scrub) and bare soil or sand.

This above-mentioned contribution difference of OSM (GUF, NTL) to the specific LCZs when being used with Sentinel-2 and Landsat-8 images reflects the difference of Sentinel-2 and Landsat-8 images, since the OSM (GUF, NTL) is the identical dataset when being used with Sentinel-2 and Landsat-8 images. Besides, the F-score difference between *S_0* and *L_0* in Figure 5 also reflects directly the difference between the two datasets. This may be related to the difference in the spectral reflectance and spatial resolution of these two datasets, as well as the detailed distinguishability required by the LCZ scheme. On the other hand, from Table 2, it can also be seen that comparable overall performance can be achieved for both Sentinel-2 and Landsat-8. This indicates the complementary information contained in Sentinel-2 and Landsat-8 images. Unfortunately, a simple stacking of both datasets together does not provide improvement, as can be seen by comparing the accuracy from *S_0*, *L_0* and *S_0 + L_0*, in Table 2. This motivates us to propose a better framework in Section 4.

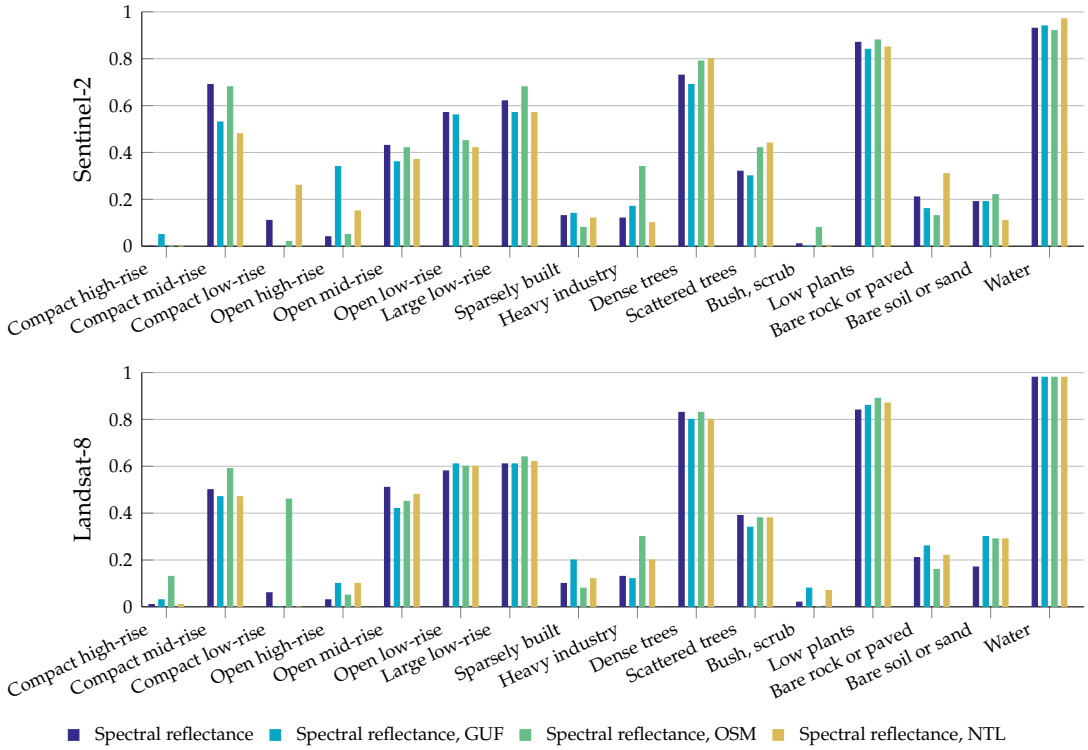


Figure 5. F-score of different classes corresponding to different feature inputs explained in Table 1.

Furthermore, it should be noticed that the contribution of the external auxiliary datasets mainly comes from the LCZs with fewer samples: compact high-rise, compact low-rise, open high-rise, sparsely built, heavy industry, bush, scrub, bare rock or paved and bare soil or sand, and spectral reflectance (the blue bars) already provides a competitive accuracy for big LCZs: compact middle-rise, open middle-rise, open low-rise, large low-rise, dense trees, scattered trees, low plants and water. This is true for both Sentinel-2 and Landsat-8 images. This can also be seen from Figure 6, which shows the relation between the F-score difference and the LCZ sample number. The F-score difference describes the difference between the highest F-score from all six configurations (*S_0*, *S_1*, *S_2*, *S_3*, *S_4*, *S_5* or *L_0*, *L_1*, *L_2*, *L_3*, *L_4*, *L_5* in Table 2) and that from configuration *S_0* or *L_0*. It shows that LCZs with more samples rely less on the external auxiliary datasets. Furthermore, it is noticed that there is an obvious class imbalance problem in LCZ classification, which is not surprising as the 17 LCZs in the real world are not comparable in quantity. We will further discuss the possible solutions for this problem, in Section 5.

Based on the above analysis, it can be concluded that the spectral reflectances of Sentinel-2 and Landsat-8 are both valuable for large-scale LCZ classification. Furthermore, no external GUF and OSM are needed when enough samples are available for all LCZs, which should be considered especially for large-scale LCZ classification, because OSM data are not available everywhere, and some cities have seen significant development since the available GUF was produced.

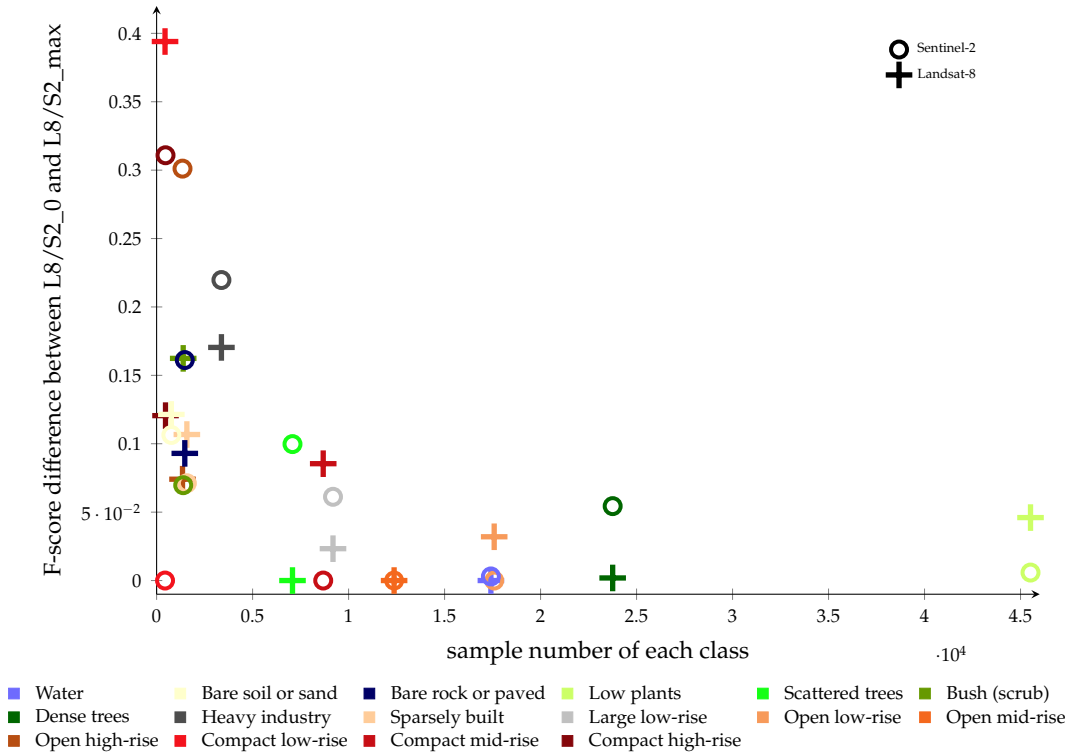


Figure 6. The relation between F-score difference and the sample number of each class.

4. Improving LCZ Classification Accuracy with Proper Input Configurations

Based on the findings of the importance of Sentinel-2 and Landsat-8 imagery, as well as the extracted spectral index measures and the external auxiliary datasets, a further classification framework is proposed by applying majority voting on results from Sentinel-2 and Landsat-8 reflectance without external auxiliary datasets, in order to explore the joint benefits of both datasets and achieve better accuracy.

The framework can be explained with an example: When Berlin is used as the test city, all the other eight cities are used for training. For both Sentinel-2 and Landsat-8 imagery, we split all training samples into ten parts class-wisely. Ten sub-classifiers are trained using ten sub-datasets for both Sentinel-2 and Landsat-8 imagery, and each sub-dataset contains 90 percent (nine parts) of all the training samples; for each sub-dataset, a different 10 percent (one part) is left out. Therefore, altogether, there are 20 sub-classifiers, with ten each from Sentinel-2 and Landsat-8, respectively. In this way, the diversity of training samples is increased.

The comparative accuracy can be seen in Table 3, where the baseline accuracy of *S2_0* and *L8_0* from Table 2 is also shown. The detailed classification accuracy of the proposed framework of each test city can be seen in Table 4. The accuracy improvement of the proposed framework of each class, averaged over all nine cities, can be seen in Figure 7.

Table 3. Improved classification accuracy by applying majority voting on results from Sentinel-2 and Landsat-8. The results are averaged over 9 test cities.

	Data, Method	OA	WA	AA	Kappa
S2	all samples used (*S2_0*)	0.71	0.93	0.46	0.65
	majority voting on 10 sub-classifiers	0.72	0.94	0.51	0.65
L8	all samples used (*L8_0*)	0.72	0.93	0.48	0.67
	majority voting on 10 sub-classifiers	0.75	0.94	0.45	0.70
S2 + L8	majority voting on 20 sub-classifiers	0.78	0.95	0.51	0.73

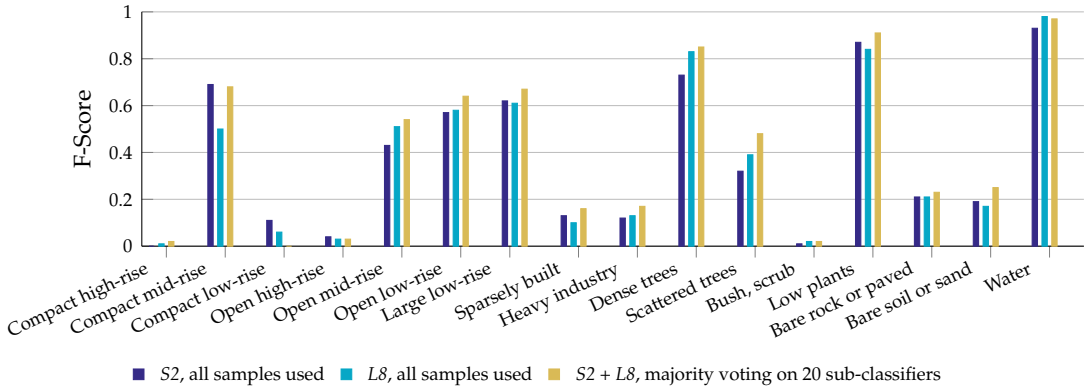


Figure 7. F-score of different classes resulting from the proposed framework and the baseline method.

The produced LCZ maps using the configuration *S2* + *L8*(in Table 3) are shown in Figure 8.

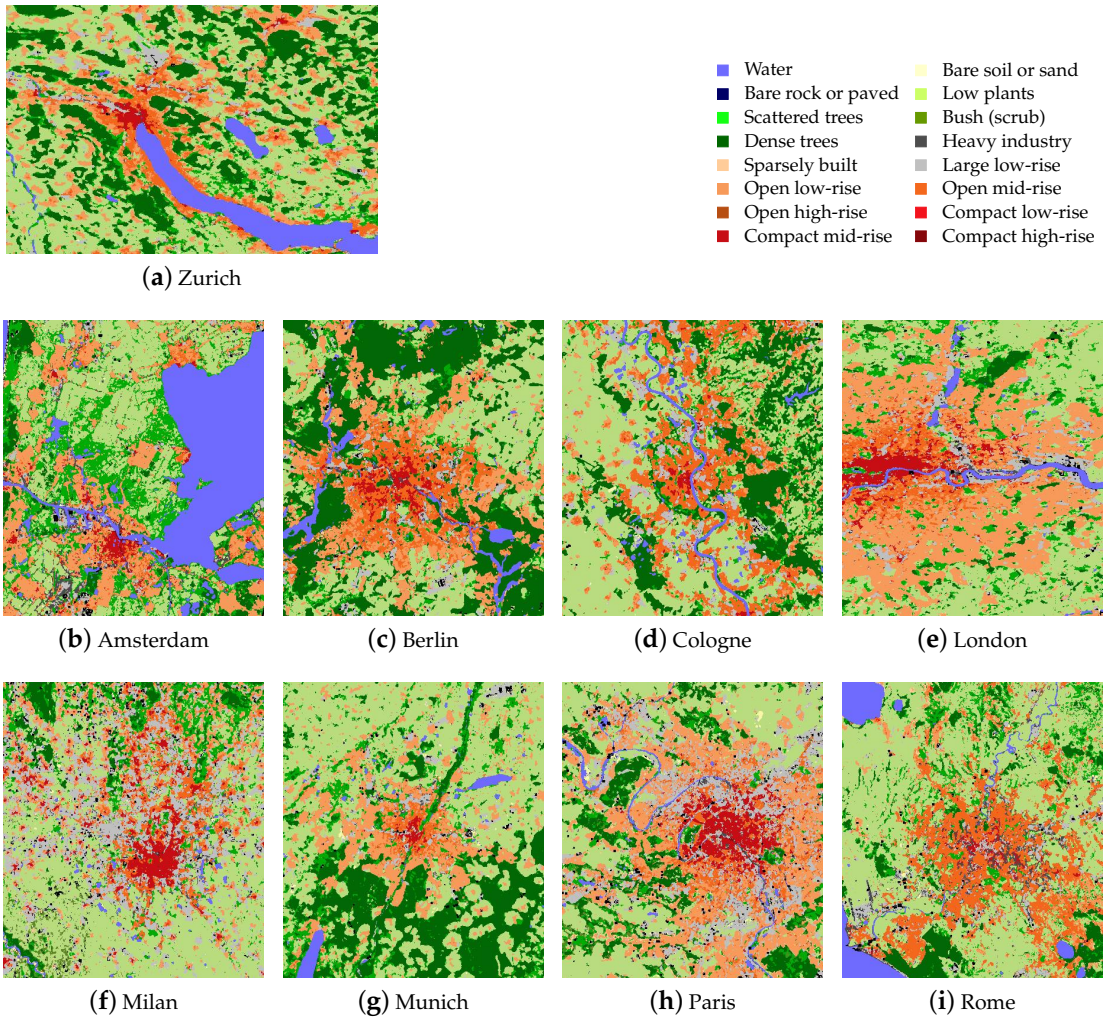


Figure 8. The LCZ maps of the nine test cities. For each city, the LCZ map is produced with the classifier trained using the reference data from the other eight cities. Only the city center is shown for each city, in order to keep a comparative size for different cities.

Table 4. Classification accuracy of each city by applying majority voting on results from Sentinel-2 and Landsat-8.

City	OA	WA	AA	Kappa
Amsterdam	0.65	0.92	0.47	0.55
Berlin	0.76	0.96	0.54	0.72
Cologne	0.78	0.96	0.49	0.73
London	0.80	0.95	0.53	0.76
Milan	0.83	0.96	0.50	0.80
Munich	0.88	0.97	0.57	0.85
Paris	0.82	0.96	0.38	0.76
Rome	0.62	0.92	0.45	0.56
Zurich	0.85	0.96	0.64	0.80
MEAN	**0.78**	**0.95**	**0.51**	**0.73**

5. Discussion

Based on the results in Section 4, the choice of datasets and the remaining challenges for large-scale LCZ mapping will be discussed in this section.

5.1. Datasets and Feature Choice for LCZ Classification

The accuracy improvement achieved by jointly using Sentinel-2 and Landsat-8, as shown in Table 3, supports the assumption in Section 3 that the two datasets contain complementary information for LCZ classification. OA has been improved from 0.72–0.78 and from 0.75–0.78, when using both datasets, compared to only using Sentinel-2 and Landsat-8 images, respectively. Improvement also exists for WA, AA and kappa. Besides, the classification accuracy is comparable among all nine test cities, as can be seen from Table 4. Furthermore, from Figure 7, we can see that 12 of all LCZs show an improvement after jointly using Sentinel-2 and Landsat-8 images.

Considering the analysis on the extracted index measures and the external auxiliary in Section 3 together, it is suggested that we should jointly use the reflectance of Sentinel-2 and Landsat-8 images for large-scale LCZ mapping, and OSM can be considered according to its availability in the specific study area. Development of more sophisticated methods to fuse the Landsat-8 and Sentinel-2 images, or even Sentinel-1, can be an interesting direction for future work.

5.2. Class Imbalance Effect

Figure 7 also shows that the accuracies of different LCZs are quite different, from lower than 10% for compact high-rise to all most 100% for water. First of all, this is due to the class imbalance problem, as mentioned in Section 3. Among the LCZs, there some are big classes, i.e., classes with an abundance of labeled samples: compact low-rise, open mid-rise, open low-rise, large low-rise, dense trees, scattered trees, low plants, water. However, there are also several small classes lacking the same amount of labeled samples: compact high-rise, compact low-rise, open high-rise, sparsely built, heavy industry, bush (scrub), bare rock or paved, bare soil or sand. It seems that balancing the training samples is necessary for LCZ mapping, as also investigated and shown by our previous work with a Canonical Correlation Forest (CCF) as the classifier [36]. Using the reflectance of Sentinel-2 over the same study area, Figure 9 illustrates the effect of different balancing methods for each LCZ. It shows that the balancing does not improve the accuracy of the big classes, while for small classes, several balancing methods improve the accuracy. However, of all the four exemplary methods, no one performs the best for all LCZs. More detailed conclusions need more systematic investigations in the future.

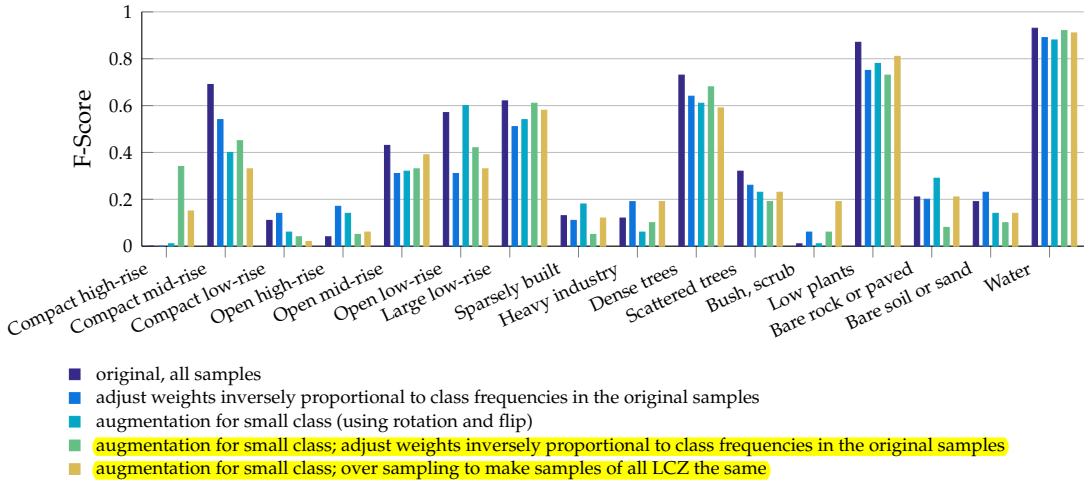
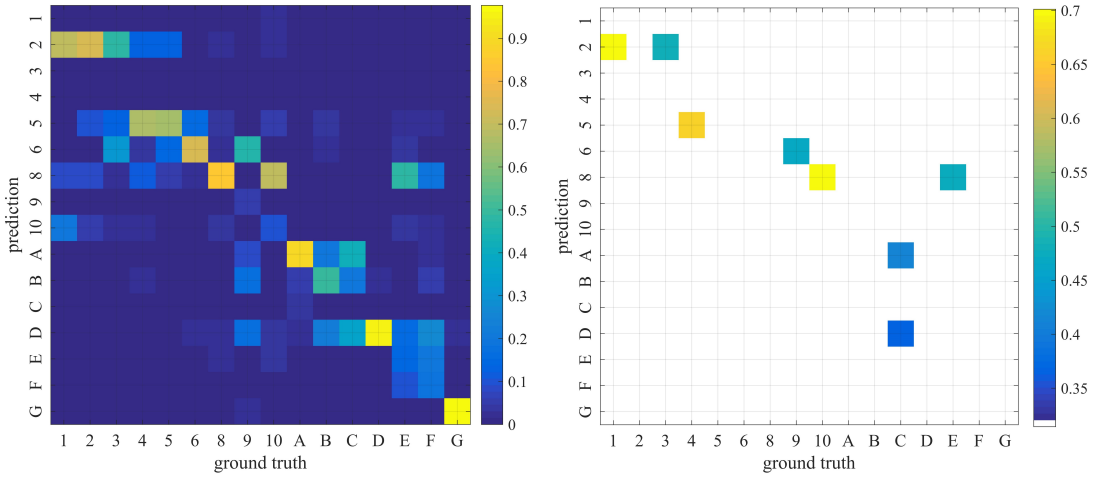


Figure 9. F-score of different classes resulting from different balancing approaches.

5.3. Confusion among LCZs

Nevertheless, even after majority voting, the AA is still less than 50% for about four of the nine test cities. The misclassification between classes can be analyzed using confusion matrices of the classification results. For conciseness, Figure 10a depicts the combined confusion matrix of all nine test cases, and Figure 10b highlights the misclassification errors higher than 30%.



(**a**) (**b**)

Figure 10. Combined confusion matrix of nine cities (**a**) and the cases with a misclassification error higher than 30% (**b**).

From Figure 10b, we can see that Class 1 (compact high-rise) and Class 3 (compact low-rise) are both falsely classified into Class 2 (compact mid-rise). Class 4 (open high-rise) is falsely classified into Class 5 (open mid-rise). This kind of misclassification resulted from the challenge of distinguishing height difference using optical satellite images, since high rise, mid-rise and low rise are quite similar in the two-dimensional optical images.

The other kind of misclassification is due to inter-class similarity: Classes 9 (sparsely built) and 10 (heavy industry) are falsely classified into Class 6 (open low-rise) and 8 (large low-rise), respectively; Class E (bare rock or paved) is falsely classified into Class 8 (large low-rise); Class C (bush, scrub) is falsely classified into Classes A (dense trees) and D (low plants), as they appear quite similar. This is illustrated in Figure 1.

To solve these problems, one possible solution is to include additional datasets such as Synthetic Aperture Radar (SAR) images to make use of radar's unique range measurements. Another solution is to adapt the LCZ scheme considering the feasibility of optical images, or a multi-level classification might be beneficial. Furthermore, multi-temporal information contained in the multi-spectral satellite images may be exploited to improve the LCZ mapping accuracy, using the state-of-the-art recurrent convolutional neural network, as shown by [37,38]. Last but not least, negative human influence on the ground truth should be weakened to guarantee the quality of the training samples across cities [35].

On the other hand, fortunately, some kinds of misclassification among LCZs are both technically understandable and acceptable for applications, since structurally similar classes such as LCZs 2 (compact mid-rise) and 5 (open mid-rise) show also similar temperature conditions [7]. Besides, [5] reported a unneglected intra-LCZ temperature variability, which is possibly due to the intra-LCZ variation of urban structures and microscale heterogeneity of the surroundings of an LCZ. Therefore, LCZ sub-classes are suggested to be used in order to exploit the full potential of the LCZ concept with respect to intra-urban distinction of local-scale environments, which is also suggested by [8], especially in high-density cities. A similar necessary adjustment regarding the LCZ scheme is also suggested by [11] in arid cities.

6. Summary and Conclusions

This paper presents an investigation of the applicability and importance of the datasets and features for LCZ classification, focusing on the globally available Sentinel-2 and Landsat-8 imagery. Investigated features include spectral reflectance, index measures and the external auxiliary datasets (GUF, OSM layers and NTL). Using ResNet, comparative experimental analysis was carried out in a large-scale study area across nine cities in central Europe. Results based on the cross-validation show that OSM and NTL can contribute to the overall classification performance, by mainly improving the classification accuracy of some of the LCZs, while GUF does not offer big benefits for the employed ResNet. Besides, comparable classification accuracy can be achieved from Sentinel-2 and Landsat-8 images, even though they display a different contribution to different LCZs.

Regarding the data and feature choice for LCZ mapping, the spectral reflectances of Sentinel-2 and Landsat-8 together are suggested to be the input features for large-scale LCZ mapping. While we were able to prove that LCZ mapping can generally benefit from jointly using both datasets in a simple majority voting framework, we see the need for further research regarding two main issues: distinguishing different building heights from optical images and class imbalance in the available samples. With these two problems solved, an even higher classification accuracy can be achieved for this detailed classification scheme, providing accurate morphological information about cities worldwide.

Author Contributions: Conceptualization, C.Q., M.S. and X.X.Z.; Methodology, C.Q. and L.M.; Software, C.Q. and L.M.; Data Curation and Investigation, C.Q. and M.S.; Writing, C.Q. and M.S.; Supervision, M.S., P.G., and X.X.Z.; Project Administration, X.X.Z.; Funding Acquisition, X.X.Z., Resources: X.X.Z..

Funding: This work is jointly supported by the China Scholarship Council (CSC), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. ERC-2016-StG-714087, Acronym: So2Sat), the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018, www.sipeo.bgu.tum.de) and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg. The work of P. Ghamisi is supported by the "High Potential Program" of Helmholtz-Zentrum Dresden-Rossendorf.

Acknowledgments: The authors would like to thank Benjamin Bechtel for support for the weighted accuracy calculation and the inspiring discussions on the LCZ scheme.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

References

1. Stewart, I.D. Local climate zones: Origins, development, and application to urban heat island studies. In Proceedings of the Annual Meeting of the American Association of Geographers, Seattle, WA, USA, 12–16 April 2011.
2. Bechtel, B.; Alexander, P.J.; Böhner, J.; Ching, J.; Conrad, O.; Feddema, J.; Mills, G.; See, L.; Stewart, I. Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 199–219. [CrossRef]
3. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [CrossRef]
4. Stewart, I.D.; Oke, T.R.; Krayenhoff, E.S. Evaluation of the 'local climate zone'scheme using temperature observations and model simulations. *Int. J. Climatol.* **2014**, *34*, 1062–1080. [CrossRef]
5. Fenner, D.; Meier, F.; Bechtel, B.; Otto, M.; Scherer, D. Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. *Meteorol. Z.* **2017**, *26*, 525–547. [CrossRef]
6. Quan, S.J.; Dutt, F.; Woodworth, E.; Yamagata, Y.; Yang, P.P.J. Local Climate Zone Mapping for Energy Resilience: A Fine-grained and 3D Approach. *Energy Procedia* **2017**, *105*, 3777–3783. [CrossRef]
7. Quanz, J.A.; Ulrich, S.; Fenner, D.; Holtmann, A.; Eimermacher, J. Micro-scale variability of air temperature within a local climate zone in Berlin, Germany, during summer. *Climate* **2018**, *6*, 5. [CrossRef]
8. Kotharkar, R.; Bagade, A. Evaluating urban heat island in the critical local climate zones of an Indian city. *Landsc. Urban Plan.* **2018**, *169*, 92–104. [CrossRef]
9. Wicki, A.; Parlow, E. Attribution of local climate zones using a multitemporal land use/land cover classification scheme. *J. Appl. Remote Sens.* **2017**, *11*, 026001. [CrossRef]
10. Taubenböck, H.; Esch, T.; Felbier, A.; Wiesner, M.; Roth, A.; Dech, S. Monitoring urbanization in mega cities from space. *Remote Sens. Environ.* **2012**, *117*, 162–176. [CrossRef]
11. Wang, C.; Middel, A.; Myint, S.W.; Kaplan, S.; Brazel, A.J.; Lukasczyk, J. Assessing local climate zones in arid cities: The case of Phoenix, Arizona and Las Vegas, Nevada. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 59–71. [CrossRef]
12. United Nations General Assembly. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
13. Danylo, O.; See, L.; Gomez, A.; Schnabel, G.; Fritz, S. Using the LCZ framework for change detection and urban growth monitoring. In Proceedings of the 19th EGU General Assembly, EGU2017, Vienna, Austria, 23–28 April 2017; Volume 19, p. 18043.
14. Ho, H.C.; Lau, K.K.L.; Yu, R.; Wang, D.; Woo, J.; Kwok, T.C.Y.; Ng, E. 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* **2017**, *14*, 994. [CrossRef] [PubMed]
15. Yokoya, N.; Ghamisi, P.; Xia, J. Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1197–1200.
16. Xu, Y.; Ma, F.; Meng, D.; Ren, C.; Leung, Y. A co-training approach to the classification of local climate zones with multi-source data. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1209–1212.
17. Esch, T.; Marconcini, M.; Felbier, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M.; Taubenböck, H.; Müller, A.; Dech, S. Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1617–1621. [CrossRef]
18. Esch, T.; Taubenböck, H.; Roth, A.; Heldens, W.; Felbier, A.; Schmidt, M.; Mueller, A.A.; Thiel, M.; Dech, S.W. TanDEM-X mission-new perspectives for the inventory and monitoring of global settlement patterns. *J. Appl. Remote Sens.* **2012**, *6*, 061702. [CrossRef]
19. Klotz, M.; Kemper, T.; Geiß, C.; Esch, T.; Taubenböck, H. How good is the map? A multi-scale cross-comparison framework for global settlement layers: Evidence from Central Europe. *Remote Sens. Environ.* **2016**, *178*, 191–212. [CrossRef]

20. Shi, K.; Yu, B.; Huang, Y.; Hu, Y.; Yin, B.; Chen, Z.; Chen, L.; Wu, J. Evaluating the ability of NPP-VIIRS nighttime light data to estimate the gross domestic product and the electric power consumption of China at multiple scales: A comparison with DMSP-OLS data. *Remote Sens.* **2014**, *6*, 1705–1724. [CrossRef]
21. Chen, Z.; Yu, B.; Hu, Y.; Huang, C.; Shi, K.; Wu, J. Estimating house vacancy rate in metropolitan areas using NPP-VIIRS nighttime light composite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2188–2197. [CrossRef]
22. Sharma, R.C.; Tateishi, R.; Hara, K.; Gharechelou, S.; Iizuka, K. Global mapping of urban built-up areas of year 2014 by combining MODIS multispectral data with VIIRS nighttime light data. *Int. J. Digit. Earth* **2016**, *9*, 1004–1020. [CrossRef]
23. Elvidge, C.D.; Baugh, K.; Zhizhin, M.; Hsu, F.C.; Ghosh, T. VIIRS night-time lights. *Int. J. Remote Sens.* **2017**, *38*, 5860–5879. [CrossRef]
24. Wang, Q.; Zhang, F.; Li, X. Optimal Clustering Framework for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [CrossRef]
25. Wang, Q.; He, X.; Li, X. Locality and Structure Regularized Low Rank Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**. [CrossRef]
26. Yokoya, N.; Ghamisi, P.; Xia, J.; Sukhanov, S.; Heremans, R.; Tankoyeu, I.; Bechtel, B.; Saux, B.L.; Moser, G.; Tuia, D. Open Data for Global Multimodal Land Use Classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1363–1377. [CrossRef]
27. Radoux, J.; Chomé, G.; Jacques, D.C.; Waldner, F.; Bellemans, N.; Matton, N.; Lamarche, C.; D'Andrimont, R.; Defourny, P. Sentinel-2's potential for sub-pixel landscape feature detection. *Remote Sens.* **2016**, *8*, 488. [CrossRef]
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
29. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
30. Zhu, X.X. So2Sat LCZ42: A Benchmark Dataset for Local Climate Zones Classification. **2018**, to appear.
31. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
32. Ji, L.; Geng, X.; Sun, K.; Zhao, Y.; Gong, P. Target detection method for water mapping using Landsat-8 OLI/TIRS imagery. *Water* **2015**, *7*, 794–817. [CrossRef]
33. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
34. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
35. Bechtel, B.; Demuzere, M.; Sismanidis, P.; Fenner, D.; Brousse, O.; Beck, C.; Van Coillie, F.; Conrad, O.; Keramitsoglou, I.; Middel, A.; et al. Quality of Crowdsourced Data on Urban Morphology—The Human Influence Experiment (HUMINEX). *Urban Sci.* **2017**, *1*, 15. [CrossRef]
36. Qiu, C.; Schmitt, M.; Ghamisi, P.; Zhu, X. Effect of the training set configuration on sentinel-2-based urban local climate zone classification. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, Riva del Garda, Italy, 4–7 June 2018; Volume 42.
37. Mou, L.; Bruzzone, L.; Zhu, X.X. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *arXiv* **2018**, arXiv:1803.02642.
38. Wang, Q.; Liu, S.; Chanussot, J.; Li, X. Scene Classification with Recurrent Attention of VHR Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *99*, 1–13. [CrossRef]