

# Fusion of Heterogeneous Earth Observation Data for the Classification of Local Climate Zones

Guichen Zhang; Pedram Ghamisi, *Senior Member, IEEE*; Xiao Xiang Zhu, *Senior Member, IEEE*

**Abstract**—*This is a preprint. To read the final version please visit IEEE XPlore.*

This paper proposes a novel framework for fusing multi-temporal, multispectral satellite images and OpenStreetMap (OSM) data for the classification of local climate zones (LCZs). Feature stacking is the most commonly-used method of data fusion but does not consider the heterogeneity of multimodal optical images and OSM data, which becomes its main drawback. The proposed framework processes two data sources separately and then combines them at the model level through two fusion models (the landuse fusion model and building fusion model), which aim to fuse optical images with landuse and buildings layers of OSM data, respectively. In addition, a new approach to detecting building incompleteness of OSM data is proposed. The proposed framework was trained and tested using data from the 2017 IEEE GRSS Data Fusion Contest, and further validated on one additional test set containing test samples which are manually labeled in Munich and New York. Experimental results have indicated that compared to the feature stacking-based baseline framework the proposed framework is effective in fusing optical images with OSM data for the classification of LCZs with high generalization capability on a large scale. The classification accuracy of the proposed framework outperforms the baseline framework by more than 6% and 2%, while testing on the test set of 2017 IEEE GRSS Data Fusion Contest and the additional test set, respectively. In addition, the proposed framework is less sensitive to spectral diversities of optical satellite images and thus achieves more stable classification performance than state-of-the-art frameworks.

**Index Terms**—Local climate zones (LCZs), heterogeneous data fusion, satellite images, OpenStreetMap (OSM), canonical correlation forest (CCF).

## I. INTRODUCTION

URBANIZATION has raised widespread concerns during the past few decades [1]–[3]. Many urban climate models

This work is jointly supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. [ERC-2016-STG-714087], Acronym: So2Sat), Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, [www.sipeo.bgu.tum.de](http://www.sipeo.bgu.tum.de)), and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg. (*Corresponding Author: Xiao Xiang Zhu*)

G. C. Zhang is with the School of Remote Sensing and Information Engineering, Wuhan University, 430079, Wuhan, Hubei, China, and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234, Wessling, Germany (email: guichen.zhang@dlr.de).

P.Ghamisi is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Weling, Germany; and is with Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology (HIF), Exploration, D-09599 Freiberg, Germany (email: p.ghamisi@gmail.com)

X. Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Germany and with Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Germany (e-mails: xiao.zhu@dlr.de).

have been formed in order to study the combined effect of urban climate and climate change on urban areas and to assess the vulnerability of urban populations [4]. It is, therefore, necessary to use a quantitative urban landscape description as the input of urban climate models [4].

Most of the studies dedicated to the urban landscape description concentrated either on separating urban areas from rural areas [5], [6] or generating local climates under different standards [4]. However, the binary schemes that separate urban areas from rural areas were not enough to characterize cities because there were many sub local climates under urban or rural categories that were nontrivial for urbanization studies [4], [7]. Consequently, a standardized scheme to characterize cities was lacking, making it hard to compare and combine their urbanization works on global and local scales [4].

Local climate zones (LCZs) is the first classification scheme providing a generic, complete, largely comprehensive, and disjoint discretization of urban landscapes with respect to the internal physical structures of urban areas on a global scale [4], [8]. The LCZ scheme is based on urban functions and climate-relevant surface properties, instead of only build-ups, which are more appropriate for urban studies [8]. Besides, it is a globally standardized and generalized scheme with inter-city comparability, and it is nonspecific to time, place, and culture [4]. The LCZ scheme describes urban areas in different levels of detail. This paper considers the LCZ scheme at level 0, where LCZs consist of 10 built labels and 7 landcover labels (see Fig. 1) [4].

Compared with field studies, satellites provide high spatial resolution images with continuous observations from space, offering a large potential in urban mapping. Moreover, OpenStreetMap (OSM) [10] data have become one of the most popular free-accessible maps (<https://www.openstreetmap.org>), providing the effective complement of satellite images [11], [12]. Furthermore, multi-source data fusion offers much potential for urban mapping. Due to the rich characteristics of natural processes and environments, it is rare for a single acquisition method to provide a complete understanding of certain phenomenon [13]–[15]. Multi-source data fusion considers the task from various points of view and then provides opportunities to view the whole picture. Therefore, this paper aims to fuse satellite images and OSM data for the classification of LCZs on a global scale. This involves three issues: classification, data fusion, and global mapping.

## A. Classification

In the past few decades, researchers have developed many effective and efficient methods for image classification. For

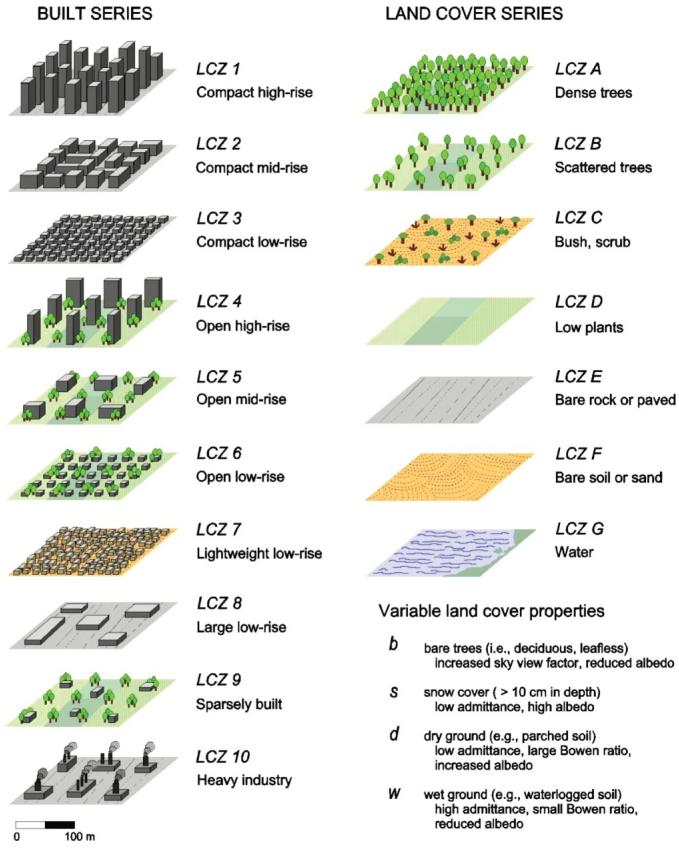


Fig. 1. LCZ Classification Scheme [9]

instance, Lu and Weng [16] gave a comprehensive review and grouped classification methods in various ways, depending on supervised or unsupervised, parametric or nonparametric, hard or soft, and pixel, sub-pixel, or object based. They summarized different classification methods as the following four points [16]. First, using supervised or unsupervised methods depends on whether training samples are available. Second, parametric methods assume that the data are subject to certain statistical distributions, which are often violated, especially in complex landscapes. Besides, much previous research has indicated that nonparametric classifiers may provide better classification results, compared to parametric classifiers, in complex landscapes. Third, hard classification assigns each pixel to a certain class, and soft classification gives a measure of belonging to each pixel. Furthermore, soft classification provides more information and potentially a more accurate result, especially for coarse spatial resolution data classification. Fourth, which level(s) of classification we use depends on the application. Pixel-level classification is straightforward and easier to implement, but it ignores the impact of mixed pixels. Sub-pixel level classification considers the heterogeneous information in one pixel and provides a more appropriate representation and area estimation of land covers than per-pixel approaches. Object-level classification firstly merges pixels into homogeneous areas and then classifies based on homogeneous areas.

This paper concentrates on the supervised, nonparametric, soft, and per-pixel classification method due to the following reasons. First, training samples are available from the contest

[17]. Second, previous works indicate that nonparametric and soft classification give a better classification performance, especially when classifying complicated urban scenes [16]. Third, for simplicity, this paper concentrates on the pixel-level classification and ignores the heterogeneous information within one pixel. In addition, we do not use object-based classification approaches, considering the application. The difference between LCZ classification and landcover mapping [18] is that LCZ labels are defined as a certain arrangement of various objects while the labels of landcover mapping are defined as objects. The segmentation process may break the certain arrangement into several pieces in the LCZ classification, where the segmented areas may lose physical structures that are key to identifying LCZ labels.

Some works [16], [19]–[21] have summarized and compared the most commonly used supervised and pixel-based classifiers, which are support vector machines (SVMs) [22], random forest (RF) [23], and neural networks (NNs) [24]. An SVM aims to find optimal linear or non-linear boundaries in high-dimensional feature spaces with or without using kernels. It is less sensitive to smaller training sets than NNs but more sensitive to the training data quality than an RF [19]. Additionally, its user-defined parameters are fewer than NNs [19]. Compared with an RF, the computation burden of an SVM is larger in the presence of a large feature quantity and when using the kernel trick [19]. An RF is an ensemble of many weak classifiers (decision trees). It is less sensitive to smaller training sets than NNs and less sensitive to the training data quality than SVMs [19]. Recently, Rainforth and Wood [25] proposed an improved forest method, called canonical correlation forest (CCF), which naturally embeds the correlation between input features and labels in hyperplane splits and outperforms 179 classifiers considered in a recent extensive survey paper [26]. Based on the studies reported in [21], a CCF outperforms an SVM, RF, and NNs in terms of classification accuracies for hyperspectral data. NNs are currently a popular method and aim to tune hyper-parameters of neural networks. It can achieve a quite good classification performance with well-determined conditions [20], [24]. An NN is usually sensitive to smaller training sets and training data quality and has severe over-fitting problems when there are not enough training samples. The classification performance highly depends on the architecture of NNs, which contains many user-defined parameters [19], [24]. This creates a very high computation burden, especially when the network goes deeper [24].

This paper aims to develop an efficient and worldwide adaptive framework for the classification of LCZs. We, therefore, intend to choose one or several classifiers with less computational burden, less over-fitting, and higher transferability among different geo-locations and better robustness over noise. As a result, in this paper, the CCF was chosen among all types of classifiers. A more detailed description of CCF will be provided in the methodology section.

### B. Data Fusion

The World Urban Database and Access Portal Tools (WUDAPT) project (<http://www.wudapt.org/>) was launched in 2012, with the aim of developing worldwide urban local climate mappings [27]. It has provided a standard classification framework that generates LCZ maps by using freely available optical satellite images, such as Landsat-8, and manually selected ground truth on Google Earth. In addition to the use of spectral bands captured by satellite images, some frameworks have also jointly considered several data sources, such as temperature [7], [28], building height [7], [29], mean amplitude of synthetic aperture radar (SAR) images [28], and OSM data [30]–[34]. Most of the above fusion studies extracted features from different data sources and then applied the feature stacking approach for data fusion as feature stacking is one of the most commonly used and fastest implemented fusion methods. Those studies assumed that classification performance improves after using more features extracted from multi-source data. Lopes, Fonte, See et al. [30] fused OSM data directly with LCZ maps from WUDAPT without using feature stacking. They manually correlated the LCZ labels and OSM feature classes and then assigned the areas with typical OSM feature classes into certain LCZ labels.

Current approaches to fuse satellite images with OSM data have several limitations aroused by data heterogeneity. First, satellite images and OSM data have different kinds of acquisition techniques. Satellite images are recorded by space observations whereas OSM data are recorded by local experts. Second, satellite images and OSM data have different data forms and spatial resolutions. Satellite images are raster data with limited resolutions (10 meters to 100 meters, in this paper) whereas OSM data are in vector format, which can be rasterized into any resolutions. The differences of data form bring many difficulties in data fusion. On one side, downsampling the OSM data into the resolution of satellite images results in the loss of much valuable information on OSM; alternatively, upsampling all satellite images will significantly increase the computational burden without adding any useful information. Third, satellite images and OSM data have different noise sources. The noise sources of satellite images come from the imaging chain (e.g., satellite platform vibration, atmosphere, etc.) whereas the noise sources of OSM data are created by the individuals recording the data, causing OSM data to contain errors or incomplete recordings. The approach to manually correlating OSM and LCZ maps [30] could somehow resolve the data heterogeneity problem, but it needs human labor to consider the correlation, which costs much time and money. The other drawback is that feature classes of OSM and LCZ labels follow different classification schemes. There is the minor possibility that some areas with certain OSM features could be directly assigned to certain LCZ labels. This demonstrates the necessity of forming novel approaches to resolve issues with heterogeneous data fusion; this is a topic deeply investigated in this paper.

### C. Global Mapping

Global LCZ mapping assists greatly in studying and comparing local climates on regional and worldwide scales. Satellite images are influenced by diverse spectral information due to complicated physical procedures in the imaging chain. This spectral diversity could decrease the classification performance, especially when analyzing multi-temporal, multi-spectral, and multi-location classification [35]. Many studies have successfully generated LCZ maps of one city by labeling samples of that city, and they have achieved satisfactory classification performance (e.g., overall accuracies (OAs) were beyond 80%). Meanwhile, it is of great interest to train models from the samples of some cities and apply the models to other cities since it costs much time and human labor to label all cities worldwide. One study [28] tried to select training samples from one city for the classification of another city by using an RF. The classification accuracies dropped to 18.2%, which indicates that the knowledge transferability between different cities should be carefully considered.

Thanks to the 2017 IEEE GRSS Data Fusion Contest [17] organized by IEEE Geoscience and Remote Sensing Society, some promising works have been accomplished in multi-model remote sensing data fusion and their transferability studies in the application of LCZ classification. The contest provided training samples from five cities and test samples from four other cities. Four novel frameworks with achieved top classification accuracies were selected; their works are quite promising and intriguing. Yokoya et al. [31] introduced CCFs [25] in analyzing knowledge transferability between cities and received the best result (OA was 74.94 %) in the contest. A CCF is an advanced forest classifier that naturally incorporates both the labels and the correlation between the input features in the choice of projection for computing decision boundaries in the projected feature space. Results demonstrate that the CCF has much better performance than other forest classifiers, such as the RF [23] and rotation forest (RoF) [36], when the training and test samples are not from the same domain [37]. Besides CCFs, three more works have managed to approach the intercity transferability problem by developing a co-training process [32], ensembling various classifiers [33], and conducting object-based classification [34] approaches. The best OAs that they have achieved are 73.2%, 72.63%, and 72.38%, respectively. However, the spectral diversity between training and test samples still plays a large impact on those frameworks. Therefore, further studies are needed on creating a generalized LCZ classification framework with more stable behavior.

Accordingly, this paper proposes a novel framework of fusing satellite images and OSM data for the classification of LCZs on a global scale. First, we extract features from satellite images and OSM data, respectively, and analyze these two features separately. Second, we apply different models to the extracted features from the satellite images and OSM data instead of applying a simple stacking of those two features. In addition, we propose a simple yet effective approach to detect the areas with incomplete recordings in the OSM data. Finally, we fuse the results from different models by conducting a

weighting process. The main contributions of this paper are thus as follows:

- 1) The proposed framework analyzes the heterogeneity between satellite images and OSM data, and we conclude that current frameworks based on feature stacking have many limitations.
- 2) This paper proposes a novel idea of fusing satellite images and OSM data by taking the data heterogeneity into account. In this context, instead of simply stacking the heterogeneous features, we apply different models to various data modality [38] in a separate manner and then conduct a novel fusion approach.
- 3) The proposed fusion approaches achieve a robust classification performance on a global scale by carefully fusing OSM with satellite images.
- 4) We propose a novel approach to detect building data incompleteness by considering the correlation between buildings and the landuse layers of OSM.

The remaining portion of this paper is organized as follows. In Section II, we introduce the dataset, study regions, and data preprocessing. In Section III, we introduce the proposed framework. In Section IV, we define the baseline framework and compare its classification performance with the proposed framework. In addition, we also present the feature importance rankings and the effectiveness of the approach of detecting incomplete building recordings. Finally, we conclude our work and give future directions in Section V.

## II. DATASET

### A. Data Fusion Contest Dataset

The dataset of the data fusion contest (DFC)<sup>1</sup> identified as "grss\_dfc\_2017" [39] was made freely available by the 2017 IEEE GRSS Data Fusion Contest [17]. A detailed data description can be found in [40]. The data consist of training samples selected from five cities (Berlin, Hong Kong, Paris, Rome, and Sao Paulo) and test samples selected from four cities (Amsterdam, Chicago, Madrid, and Xi'an) (see Fig. 2). In each city, the data contain multi-temporal Landsat-8 images, single-temporal Sentinel-2 images, and OSM data. Satellite images are of 1C-level and have 100 meter spatial resolution. OSM data include buildings, landuse, water, and natural layers, which are available in both raster and vector data formats. The raster form of OSM data is of 5 meter spatial resolution. A buildings layer is a binary layer delineating building areas. A landuse layer separates an area into different landuse classes [41].<sup>2</sup> OSM data also include road layers, which are only available in the vector data form. Moreover, the distribution of training and test sizes are provided in Table I.

In this paper, we redownloaded Landsat-8 images from the U.S. Geological Survey (<https://earthexplorer.usgs.gov/>) and Sentinel-2 images from the Sentinel Data Hub

<sup>1</sup>The data fusion contest (DFC) dataset refers to 2017 IEEE GRSS Data Fusion Contest [17] unless otherwise noted.

<sup>2</sup>Landuse layers may contain the following classes: forest, park, residential, industrial, farm, cemetery, allotments, meadow, commercial, nature\_reserve, recreation\_ground, retail, military, quarry, orchard, vineyard, scrub, grass, heath, and national\_park.

(<https://scihub.copernicus.eu/>) to acquire the original spatial resolution images. Then, atmospheric corrections were conducted and cloud masks were generated. We only kept those redownloaded images with exactly the same geo-location and time acquisition as grss\_dfc\_2017 [39]. In this work, buildings and landuse layers from OSM data were only used because it was discovered that the water and natural layers were not available for all cities. In addition, road layers were not considered in this work.

Data preprocessing was conducted on Landsat-8 images, Sentinel-2 images, the buildings layers of OSM, and the landuse layers of OSM. For Landsat-8 images, atmospheric corrections using ATCOR-2/3 version 9.0.0 with the haze removal option were conducted, and then the data were upsampled into 10 meter spatial resolution using bicubic interpolation. For Sentinel-2 images, atmospheric corrections were conducted using Sen2Cor version 2.3.1, and then the data were upsampled into 10 meter spatial resolution using bicubic interpolation. After preprocessing, Landsat-8 images contained the Bands 1, 2, 3, 4, 5, 6, 7, 8, 10, and 11, and Sentinel-2 images contained the Bands 2, 3, 4, 5, 6, 7, 8, 8A, 11, and 12. We did not consider cirrus and water vapor bands because they are mainly dedicated to cirrus detections and water vapor corrections and are not usually used in urban mapping [42]. We also did not consider coastal/aerosol bands of Sentinel-2 because they are dedicated to aerosol retrieval and cloud detection [42]. Besides, a cloud mask was generated from satellite images at each acquisition time. The areas that contained high cloud probability were removed in the following process. For buildings and landuse layers of OSM data in raster form, each layer was normalized between 0 and 1. For the buildings layers of OSM data in vector form, the layers of building central points were extracted using ArcMap version 10.5.1, and then the layers of building central points were rasterized into 5 meter spatial resolution.

### B. Additional Test Dataset

In addition to the four test cities available in grss\_dfc\_2017 [39], we have used an additional test (AT) dataset by selecting two extra test cities (Munich and New York, see Fig. 2) to validate our proposed framework. Ground truth data were labeled according to the LCZ classification scheme [9]. Table I shows the distribution of test sizes. In each city, we have downloaded multi-temporal Sentinel-2 images from the Sentinel Data Hub (<https://scihub.copernicus.eu/>) and OSM data from Geofabrik (<https://www.geofabrik.de/>). After that, Sentinel-2 images and OSM data were processed according to the strategy in II.A.

## III. METHODOLOGY

In this section, we propose a novel framework to fuse satellite images with OSM data (see Fig. 3). First, spectral, spatial, textural, and map features were extracted from satellite images and OSM data. Second, three different models were applied to these three kinds of extracted features. Specifically, CCFs [25] were applied to the satellite features, and then a landuse fusion model and building fusion model were derived to fuse landuse features and building density features with satellite

TABLE I  
TRAINING AND TEST SAMPLES

Lable No.	Training Size	Test Size (DFC)	Test Size (AT)
1	1642	242	526
2	6103	4892	1414
3	5738	1535	2825
4	2098	2270	172
5	4759	2255	1131
6	8891	8265	4271
7	0	0	0
8	4889	11230	2304
9	1156	1072	3763
10	449	920	887
11	17716	3170	2175
12	2819	4528	362
13	1741	1284	41
14	14457	12994	2949
15	323	1104	253
16	503	391	0
17	8561	4454	12626



Fig. 2. Training cities (Berlin, Hong Kong, Paris, Rome, and Sao Paulo) marked with black dots from grss\_dfc\_2017 [39], test cities (Amsterdam, Chicago, Madrid, and Xi'an) marked with red dots from grss\_dfc\_2017 [39], and additional test cities (Munich and New York) marked with green dots from additional test dataset.

features. In addition, a novel approach was also proposed to mask out incomplete building areas. Finally, postprocessing and decision fusion were conducted.

#### A. Feature Extraction

Spectral, spatial, and texture features were extracted from the preprocessed Landsat-8 and Sentinel-2 images through the same computation process at each acquisition time. First, mean values and standard deviations of all bands of images in each patch of 100 m ground sample distance (GSD) were computed. Second, three spectral indexes (normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and bare soil index (BSI)) were derived, and then their mean values and standard deviations were computed in each patch of 100 m GSD. Third, the mean values of morphological profiles (MPs) of NDVI were computed in each patch of 100 m GSD. Fourth, a weighted gray-level co-occurrence matrix (GLCM) algorithm [43] was used to produce contrast, correlation, energy, and homogeneity texture

TABLE II  
EXTRACTED FEATURES

Feature Name	Quantity	Spatial Resolution
Mean values of bands	10	100
Mean values of NDVI, NDWI and BSI	3	100
Std values of bands	10	100
Std values of NDVI, NDWI and BSI	3	100
Contrast, Correlation, Energy and Homogeneity	4	100
MPs of NDVI	6	100
Building density	1	100
Building	1	5
Landuse	1	5
All	39	

features, and then their mean values were computed in each patch of 100 m GSD. Besides, building density features were extracted from the layers of building central points by counting the building number in each patch of 100 m GSD. In addition, the preprocessed landuse and buildings layers were also used as the landuse and building features.

Table II lists the extracted features' names, quantities, and spatial resolutions. Since the feature extraction from Landsat-8 or Sentinel-2 images shares the same computation process and those extracted features from the two satellites have the same feature names, the spectral, spatial, and textural features in Table II represent the extracted features from either Landsat-8 or Sentinel-2 images. Moreover, the spectral, spatial, and textural features from either Landsat-8 or Sentinel-2 images are named as satellite features.

#### B. Canonical Correlation Forest (CCF)

CCF [25] is an ensemble model based on oblique decision trees. Compared with an RF, which computes the hyperplane splits in the coordinate system of input features, a CCF naturally constructs a projected feature space by considering the correlation between input features and their corresponding labels [25]. It is more robust to the rotation, translation, and global scaling of the input features [25]. One CCF model is composed of many sub-models named canonical correlation trees (CCTs). One CCT is a binary decision tree with many sequential divisions and is regarded as the smallest predictive unit in this paper. Each CCT is trained independently, and the ensemble of CCTs can simultaneously improve predictive performance and provide regularization against over-fitting [25]. This section applies a CCF to the stacked satellite features (see Fig. 4) and then generates an initial classification result of a test city. The classification result from one CCF could be integrated into a *votes* cube. The first two dimensions are the spatial dimensions (row and column directions), which have the corresponding pixel coordinates for satellite features of the test city. The third dimension has the same length with LCZ labels, which is 17. *votes*(*i*, *j*, *l*) record the votes number of the *l*-the label in the pixel coordinate (*i*, *j*) of a test city.

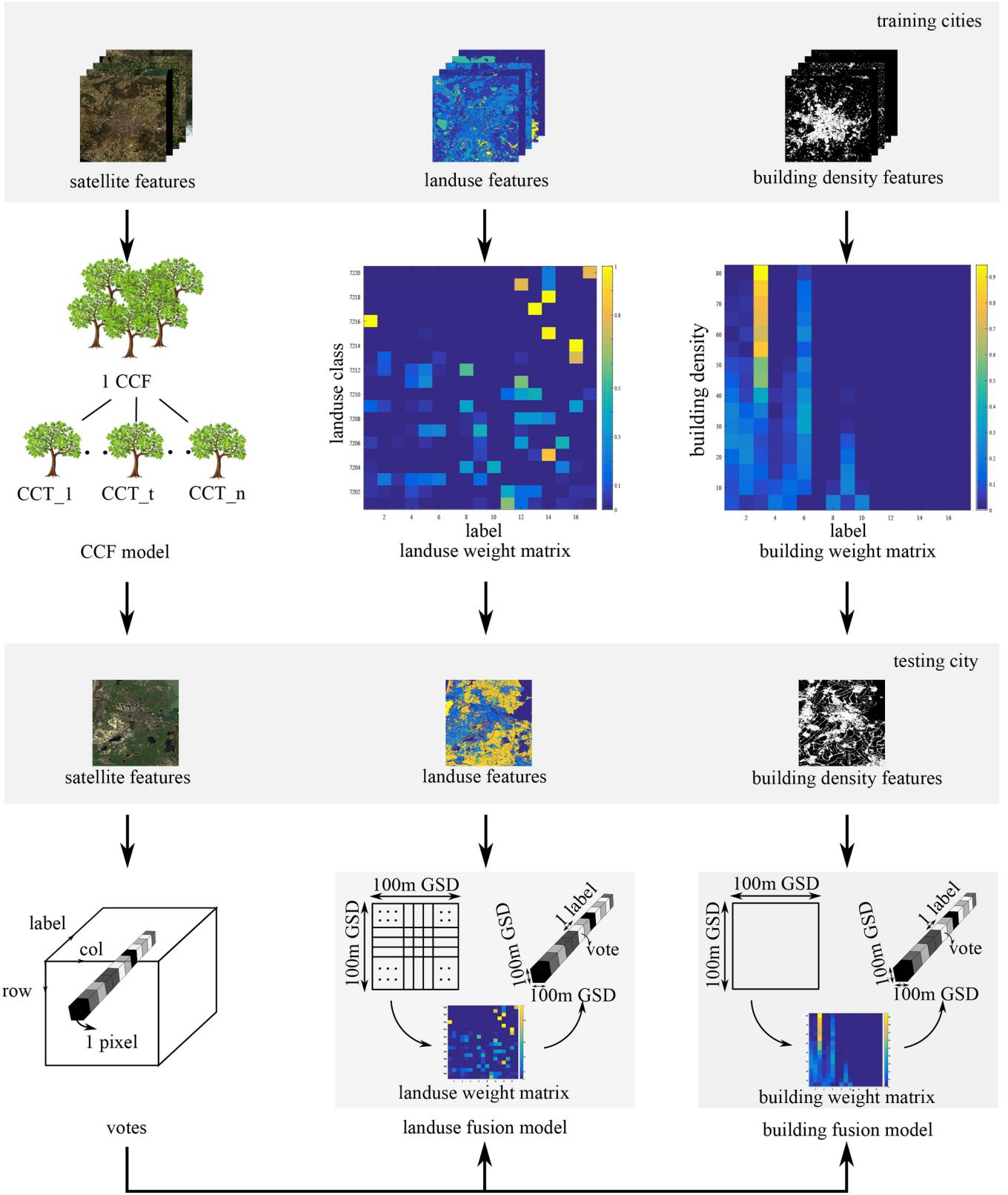


Fig. 3. Proposed framework.

The larger the votes number is, the more convincing that the pixel belongs to that label.

### C. Landuse Fusion Model

A landuse fusion model contains two parts. In the first part, the relation between landuse classes and ground truth data is

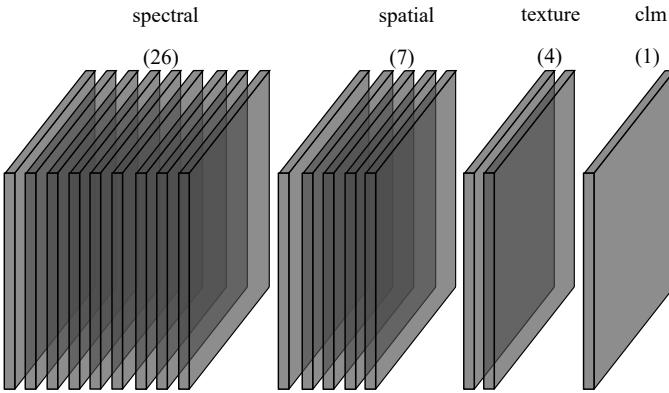


Fig. 4. Feature stacking of satellite features.

trained. In the second part, this relation is used as an active aid to fuse landuse features with satellite features.

A landuse feature is denoted as  $landuse(i, j) = lu_{\cdot}(i, j)$ , which represents the pixel coordinates.  $lu_{\cdot}$  records the value of the pixel  $(i, j)$ . Those pixels where  $landuse(i, j) = 0$  were removed in advance. First, landuse features and ground truth data from the training set were used to compute the prior knowledge, which was a 2D probability distribution matrix named the landuse weight matrix. The row direction of the matrix identifies different landuse classes, and the column direction identifies LCZ labels. Each element in the matrix records how large the probability is that each landuse class belongs to each LCZ label. The sum of each row is equal to 1. This matrix is denoted as  $lu\_wn(lu, label)$ , where  $lu$  represents different landuse classes and  $label$  represents different LCZ labels.  $lu\_wn(lu, label)$  is then fused with the  $votes$  cube generated from the CCF by using a weighting procedure. For the landuse feature  $landuse$  of a test city and its corresponding  $votes$  cube, the weighting procedure was conducted in each patch of 100 m GSD. Since the spatial resolution of  $landuse$  is 5 meters and the spatial resolution of  $votes$  is 100 meters, one pixel in  $votes$  corresponds to  $20 \times 20$  pixels in  $landuse$ . For each 100 m GSD, the weighting procedure was conducted using the formula (1):

$$luwn\_votes(i, j, :) = \sum_{d=1}^{400} lu\_wn(lu(d), :) \cdot votes(i, j, :), \quad (1)$$

where  $lu(d) \neq 0$ ,  $(i, j)$  are the pixel coordinates in the spatial domain of  $votes$ , and  $luwn\_votes(i, j, :)$  records the weighted  $votes$  vector of pixel  $(i, j)$ .

#### D. Building Fusion Model

The building fusion model also contains two parts. In the first part, the relation between building density values and ground truth data is trained. In the second part, this relation is used as an active aid to fuse building density features with satellite features.

A buildings layer is denoted as  $build(i, j) = b_{\cdot}(i, j)$  represent the pixel coordinates.  $b_{\cdot}$  records the value of the pixel

$(i, j)$  and is either 0 or 1. First, building density features and ground truth data from the training set were used to compute the prior knowledge, which was a 2D probability distribution matrix named the building weight matrix. The row direction of the matrix identifies different building density ranges, and the column direction identifies LCZ labels. Each element in the matrix records how large the probability is that each building density range belongs to each LCZ label. The sum of each row is also equal to 1. Assuming the largest building density value from training samples was  $bn\_max$ , the building density ranges  $bu$  were defined according to formula (2):

$$[0, gap], [gap+1, 2gap], \dots, [bn\_max+1, bn\_max+gap], \quad (2)$$

where  $gap$  is an empirical value equal to 5 in this paper.

We use the building density ranges, instead of each building density value, because the building density values from training samples cannot cover all values from 0 to  $bn\_max$  due to the limited number of training samples.

The building weight matrix is denoted as  $bu\_wn(bu, label)$ , where  $bu$  represents different building density ranges and  $label$  represents different LCZ labels.  $bu\_wn(bu, label)$  is then fused with the  $votes$  cube generated from the CCF by using a weighting procedure. For the building density feature  $build$  of a test city and its corresponding  $votes$  cube, the weighting procedure was conducted on each patch of 100 m GSD. Since  $build$  and  $votes$  have the same spatial resolution (i.e., 100 meters), the weighting procedure could be directly computed for each 100 m GSD according to formula (3):

$$buwn\_votes(i, j, :) = bu\_wn(bu, :) \cdot votes(i, j, :) \quad (3)$$

where  $(i, j)$  are the pixel coordinates in the spatial domain of  $votes$ , and  $buwn\_votes(i, j, :)$  records the weighted  $votes$  vector of pixel  $(i, j)$ .

One remaining problem still existed before using the building fusion model. Since the building density value is computed by counting the number of buildings in a local area, the building density value could be completely wrong if the buildings features have incomplete data recordings in that area. Therefore, it is necessary to generate building confidence masks in order to automatically mask out those areas with incomplete recordings of building data.

#### E. Building Confidence Mask Generation

This paper proposes a novel approach to compute building confidence in a local area by jointly considering landuse and building features under 5 meter spatial resolution without using satellite images (see Fig. 5). The idea is that at least one building pixel should be near the local area of the pixel, which indicates a high building probability. This approach consists of four steps that will be discussed in the following subsections.

1) *Building Probability Generation:* Landuse and building features from the OSM data contain correlated knowledge. This step generated a 2D probability distribution matrix (see Fig. 6) recording the relation between landuse and building features by using both training and test samples. Zero values in landuse features were removed in advance. The row direction

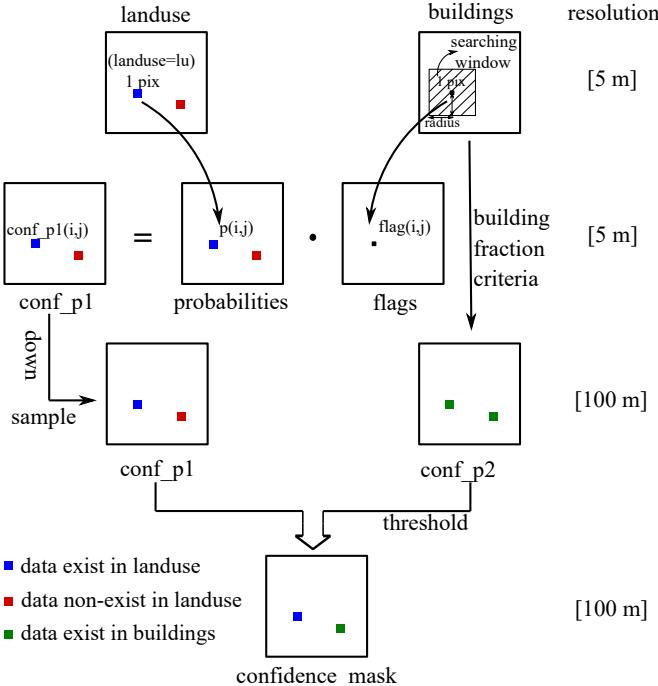


Fig. 5. Building confidence mask generation.

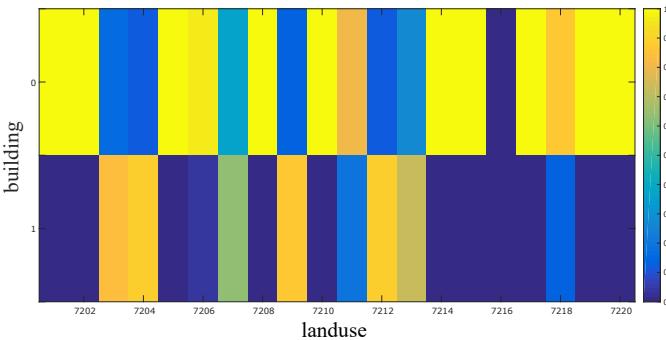


Fig. 6. Building-landuse probability distribution.

of the matrix identifies two different building classes: building and non-building. The column direction identifies different landuse classes. Each element in the matrix records how large the probability is that each landuse class belongs to each building class. The second row of this matrix is denoted as  $p(build = 1|landuse = lu)$ , which is used in this approach. Fig. 6 presents some hints about this relation. For example, the residential class in the landuse layers has a quite high probability of being a building pixel in the buildings layers whereas the forest class in the landuse layers has a quite low probability of being a building pixel in the buildings layers.

2) *Local Searching*: After generating the relation matrix, local searching was then conducted in the buildings feature of each city. The size of the searching area was given empirically after considering building intervals and is equal to 5 pixels in this paper. For each pixel in a landuse feature, we searched building pixels around that pixel in the corresponding building feature. We used a binary value  $flag$  to record the searching result. The building confidence  $conf\_p1$  was computed ac-

TABLE III  
BUILDING SURFACE FRACTION OF LCZS [8]

Label	Building Surface Fraction (%)
Compact high-rise	40-60
Compact mid-rise	40-70
Compact low-rise	40-70
Open high-rise	20-40
Open mid-rise	20-40
Open low-rise	20-40
Lightweight low-rise	60-90
Large low-rise	30-50
Sparingly built	10-20
Heavy industry	20-30
Dense trees	0-10
Scattered trees	0-10
Bush, scrub	0-10
Low plants	0-10
Bare rock or paved	0-10
Bare soil or sand	0-10
Water	0-10

cording to:

$$conf\_p1 = p(build = 1|landuse = lu) \cdot flag, \quad (4)$$

where

$$flag = \begin{cases} 1, & \text{searching succeeded} \\ -1, & \text{searching failed.} \end{cases} \quad (5)$$

3) *Detection Complement*: The previous step is not applicable in those pixels where the values in landuse layers are zero. Therefore, this step, which is independent of the previous step, aims to provide supplementary information when the previous step cannot be applied. Thanks to the quantitative standard from the LCZ definition [8], each label of LCZs defines a range of building surface fractions (see Table III).

The building surface fraction was computed in each patch of 100 m GSD from building features. Then the empirical value of 10% was used to compute the building confidence value  $conf\_p2$ .  $conf\_p2 = 1$  if the building surface fraction was higher than 10%; otherwise,  $conf\_p2(i, j) = 0$ . 10% was used as the threshold for two reasons. First, 10 % is the lowest boundary of building surface fractions among all built types, so that sparsely built can be kept. Second, this parameter was not sensitive to the final result, which will be explained in detail in the experimental part.

4) *Combination*: The local searching result  $conf\_p1$  is combined with the building fraction result  $conf\_p2$  according to formula (6):

$$conf\_comb(i, j) = \begin{cases} conf\_p2(i, j), & landuse(i, j) = 0 \\ conf\_p2(i, j), & conf\_p2(i, j) = 1 \\ conf\_p1(i, j), & \text{otherwise} \end{cases} \quad (6)$$

Afterwards, we used an empirical value (e.g., 0.8) to threshold  $conf\_comb$  into a binary layer, where  $conf\_comb = 1$

identified building confident areas and  $conf\_comb = 0$  identified building unconfident areas.

#### F. Postprocessing and Decision Fusion

The above process could generate one weighted *votes* cube at each satellite acquisition time for each satellite in each test city. Therefore, if one has  $T$  acquisition times for Landsat-8 images and one acquisition time for Sentinel-2 images of one test city, then  $T + 1$  weighted *votes* cubes can be obtained. Next,  $T + 1$  classification maps could be computed by selecting the label with the largest votes number per pixel. Then, the median filter was applied with the size of [3,3] to those classification maps. Finally, decision fusion was conducted among the  $T + 1$  classification maps through majority voting.

## IV. EXPERIMENT

In this section, we firstly introduce a baseline framework and compute the contributions of the extracted features. Then, the classification performance of the proposed framework and the baseline framework is compared according to the classification accuracies and framework transferability. Afterwards, another experiment demonstrates the effectiveness of the proposed approach in generating building confidence masks.

The following parameters were set empirically in the experiments:

- 1) MPs: A disk-shaped structuring element was used whose sizes were 1,4,7, and 10.
- 2) GLCM: The number of gray levels was 32. Directions were  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . The offset that defined the distance of the spatial adjacency was 1 pixel.
- 3) CCF: The number of CCTs was 20 to follow the literature in [31].
- 4) Building fusion model: The interval of building densities in the building weight matrix was 5. The radius of the local building searching was 25 meters GSD. The threshold of the building surface fraction was 10%. The threshold of generating building confidence masks was 0.8.

#### A. Baseline Framework and Feature Importance

The baseline framework was used to evaluate the performance of the proposed classification framework. The baseline framework directly stacks the features in Table II and feeds those features into the CCF. The building and landuse features with 5 meter spatial resolution were firstly down-sampled into 100 meter spatial resolution using the "nearest neighbor" before feature stacking. Two groups were considered where OSM data were stacked with Landsat-8 and Sentinel-2 images. Therefore, the baseline and the proposed frameworks are comparable, and the difference in their classification performance was aroused by the feature fusion models. In addition, feature importance was computed by using training samples through a 5-fold cross-validation to conclude which features contributed more to the classification performance. Fig. 7 illustrates the contributions of the extracted features from Landsat-8 and

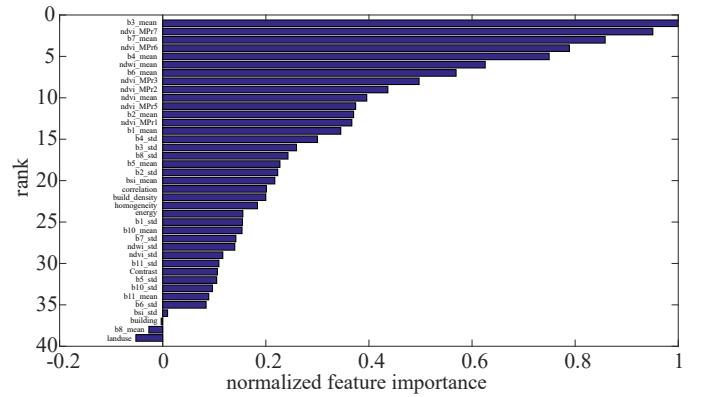


Fig. 7. Feature importance of the Landsat-8 group.

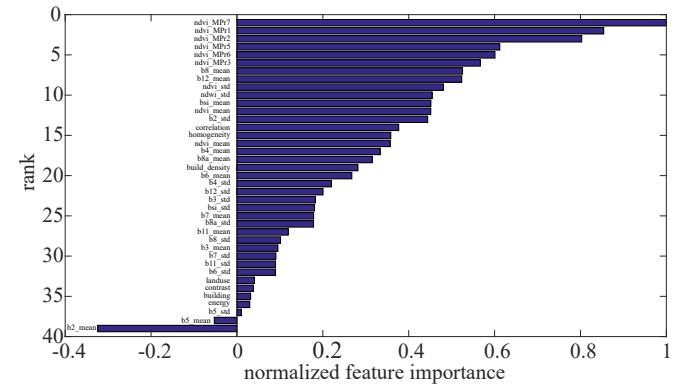


Fig. 8. Feature importance of the Sentinel-2 group.

OSM data. Fig. 8 demonstrates the contributions of the extracted features from Sentinel-2 and OSM data.

These two feature importance maps share many similar characteristics. First, NDVI and its MPs, which contain the information of vegetation abundance and its spatial information, obviously have higher importance than other features. The quantitative properties of LCZs could somehow explain why NDVI and its MPs are quite important here. Table IV illustrates the pervious and impervious surface fractions of LCZ labels, indicating that vegetation abundance is related to built types. For example, the compact high-rise, compact mid-rise, and compact low-rise have different ranges of pervious and impervious surface fractions in spite of the fact that they all belong to the compact-built. Therefore, although vegetation abundance could not make the built types completely separable, it still played an important role in separating different built types.

Second, both rankings of the buildings and landuse features were quite low. The landuse feature ranked the last place and the 33rd place in the Landsat-8 and Sentinel-2 groups, respectively. The building feature ranked the 37th place and the 35th place in the Landsat-8 and Sentinel-2 groups, respectively. These two facts indicate that landuse and building features contribute trivial importance (even negative importance) in LCZ classification if directly using feature stacking due to the data heterogeneity. Therefore, novel models are highly necessary to fuse OSM and satellite features.

TABLE IV  
PERVIOUS AND IMPERVIOUS SURFACE FRACTION OF LCZs [8]

Label	Pervious Surface Fraction (%)	Impervious Surface Fraction (%)
Compact high-rise	0-10	40-60
Compact mid-rise	0-20	30-50
Compact low-rise	0-30	20-50
Open high-rise	30-40	30-40
Open mid-rise	20-40	30-50
Open low-rise	30-60	20-50
Lightweight low-rise	0-30	0-20
Large low-rise	0-20	40-50
Sparsely built	60-80	0-20
Heavy industry	40-50	20-40
Dense trees	90-100	0-10
Scattered trees	90-100	0-10
Bush, scrub	90-100	0-10
Low plants	90-100	0-10
Bare rock or paved	0-10	90-100
Bare soil or sand	90-100	0-10
Water	90-100	0-10

Third, the building density feature ranked higher than the building feature itself, in 22nd and 19th place, respectively. This could be due to the building interval, one of the most distinct criteria of separating different LCZ built types, which is highly related to building density values in local areas. Building features that only delineate building areas do not contain spatial information and thus have trivial contribution to classification.

### B. Accuracies Improvement

1) *Data Fusion Contest Dataset*: Fig. 9 and Fig. 10 provide the classification maps generated by the proposed framework with both fusion models and the baseline framework on four test cities. Table V compares their classification accuracies in terms of overall accuracy (OA) and kappa coefficient (kappa).

First, the OA and kappa of the proposed framework were 76.15% and 0.72, which outperformed the baseline accuracy by 6.01% and 7%, respectively. Furthermore, the OA and kappa of the proposed framework still outperformed the winner of the 2017 IEEE GRSS Data Fusion Contest [17], [31], [37] by 1.21% and 1%, respectively, although fewer classifiers were used in this paper.

Second, the use of the landuse or building model could significantly improve the accuracies in general. For example, compared with the baseline framework, the OA in Amsterdam increased by 11.2% and 8.71%, respectively. The individual use of the landuse or building model may sometimes decrease the accuracies in some cities. For example, the OA decreased by 1.76% after applying the landuse model to the city of Chicago and by 0.49% after applying the building model to the city of Madrid. One reason could be that training samples are still quite limited to well represent the complexities of the test samples, especially when OSM data have many incomplete recordings. Another reason could be that landuse and building

fusion models are applied to different areas of test cities if landuse and buildings layers lack data in different areas. Consequently, combining both models could tune the pre-classification results to the largest extent. Although accuracies may occasionally decrease when using a single fusion model, the decreased values are much smaller than the increased values when considering the four test cities. Therefore, applying the landuse or building model solely to test cities still increases the accuracies in general.

Third, the joint use of both fusion models always increases the accuracies significantly and is consistently better than using a single model. For example, the OA increased by 11.2% and 8.71%, respectively, after applying the landuse and building fusion models separately to Amsterdam, but the OA increased by 20.67% after applying both models to Amsterdam at the same time. Additionally, the OA increased by 3.61% and decreased by -0.49%, respectively, after applying the landuse and building fusion models separately to Madrid, but the OA increased by 4.88% after applying both models to Madrid at the same time. This indicates that the use of both models simultaneously is not only an improvement over using each model alone but also demonstrates how the two models can aid one another and further boost the classification performance.

Fourth, landuse and building models impact the accuracies in different test cities to various degrees. For example, the OA increased by 20.67% in Amsterdam, but it only increased by 0.93% in Chicago after using both fusion models together. This may indicate that training information from only five cities is not adequate to train a generalized landuse or building fusion model, especially considering that landuse and buildings layers lack much data in certain cities, such as Xi'an, Sao Paulo, Hong Kong, etc.

Besides the OA and kappa, we also compared the improvement of the distributions of producer accuracies (PAs) (Fig. 11). First, accuracies of all labels were improved except for label 1 (compact high-rise), where several samples were incorrectly classified as label 8 (large low-rise). The possible reasons are, first, test samples of label 1 are quite few, and the building densities between label 1 and 8 could be similar. Second, the largest improvement occurred in label 12 (scattered trees), which was highly mixed with label 14 (low plants) in the baseline framework. A typical example exists in the classification map of Madrid (see Fig. 12). Compared with the proposed framework, the areas of scattered trees changed more significantly among different acquisition times in the case of the baseline framework. It indicates that the proposed framework is more robust to spectral changes, because OSM data have provided active aid to optical observation. Besides, the proposed framework also demonstrates a large amount of improvement in label 10 (heavy industry), label 6 (open low-rise), label 5 (open mid-rise), label 4 (open high-rise), and label 8 (large low-rise). For example, the PA of label 10 (heavy industry) was less than 10% through the use of the baseline framework and was highly mixed with label 8 (large low-rise), but it increased to 11% when using the proposed framework. Analyzing satellite images is considerably helpful for retrieving spectral, spatial, and textural information. Occasionally, retrieved knowledge of two areas from satellite images may

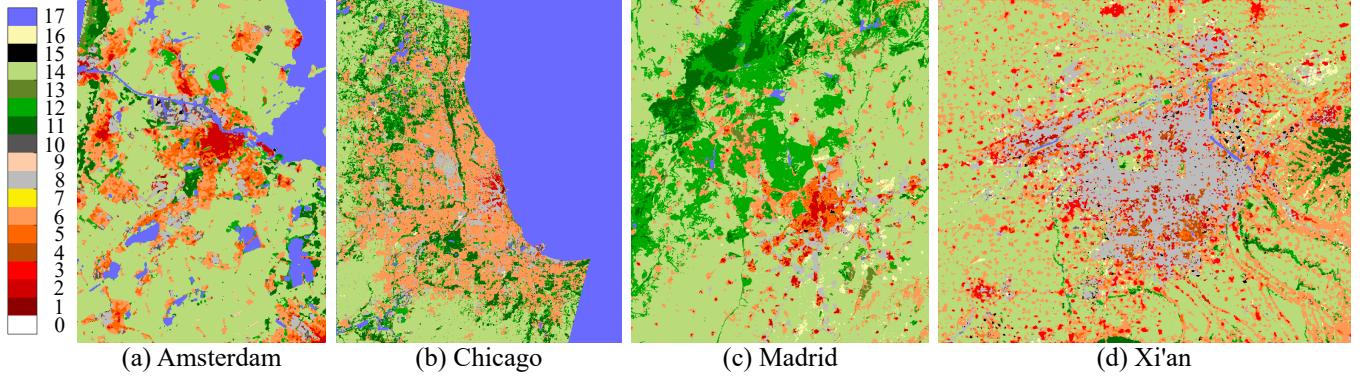


Fig. 9. Classification maps from proposed framework (DFC dataset).

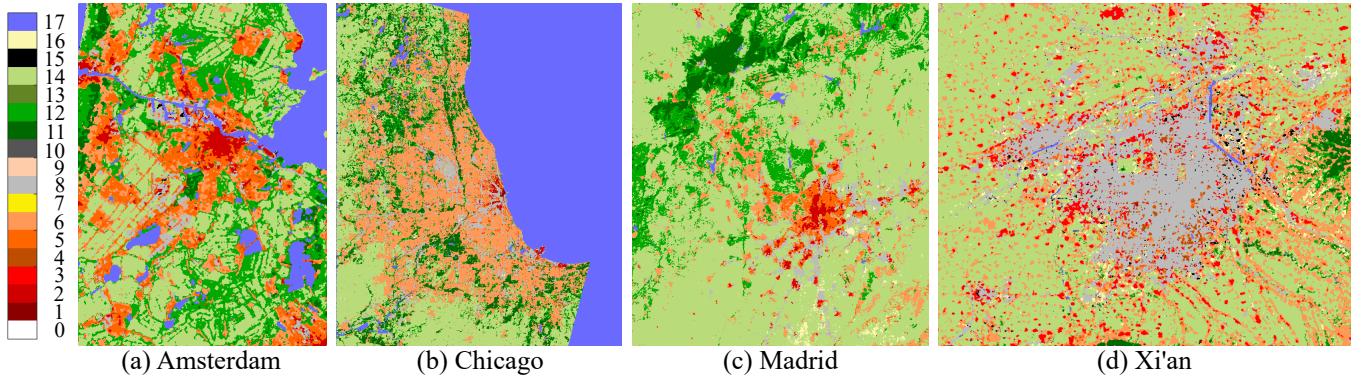


Fig. 10. Classification maps from baseline framework (the DFC dataset).

TABLE V  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK (THE DFC DATASET)

City	Baseline	Landuse Fusion Model	Building Fusion Model	Both Models
Amsterdam	OA=61.50 K=0.56	OA=72.52 K=0.68	OA=70.21 K=0.66	OA=82.17 K=0.79
	OA=77.63 K=0.73	OA=75.87 K=0.71	OA=78.65 K=0.74	OA=78.56 K=0.74
Chicago	OA=74.24 K=0.69	OA=77.85 K=0.74	OA=73.75 K=0.69	OA=79.12 K=0.75
	OA=53.33 K=0.44	OA=59.31 K=0.52	OA=53.67 K=0.45	OA=58.00 K=0.5
All	OA=70.14 K=0.65	OA=73.63 K=0.70	OA=71.41 K=0.67	OA=76.15 K=0.72

appear to be similar (labels 8 and 10); however, these two areas belong to different LCZ labels. Compared to satellite images, OSM data contain advanced knowledge (e.g., urban functionalities) that separates different classes through human intelligence, offering non-trivial assistance in the classification of LCZs.

Meanwhile, the accuracies of some labels are still not satisfactory, mainly because of the complicated scheme of LCZs. First, label 3 (compact low-rise) and label 4 (open high-rise) are still highly mixed with other built types. It is still challenging to precisely describe urban structures due to the diverse construction of different cities. Second, label 9 (sparsely built) is highly mixed with label 6 (open low-rise) and label 14 (low plants), likely due to the classification

scale. After checking ground truth data from Google Earth, it was observed that sparsely built areas often follow a "cluster" behavior. Stated another way, in sparsely built areas, buildings appear to be dense in certain sections and absent in other sections. Third, label 15 (bare rock or paved) is highly mixed with label 8 (large low-rise), likely due to training samples. After checking ground truth data from Google Earth, it was observed that paved ground frequently appears near large low-rises and is occasionally selected for training samples of large low-rises.

2) *Additional Test Dataset:* Fig. 13 and Fig. 14 provide the classification maps generated by the proposed framework with both fusion models and the baseline framework on Munich and New York. Table VI compares their classification accuracies

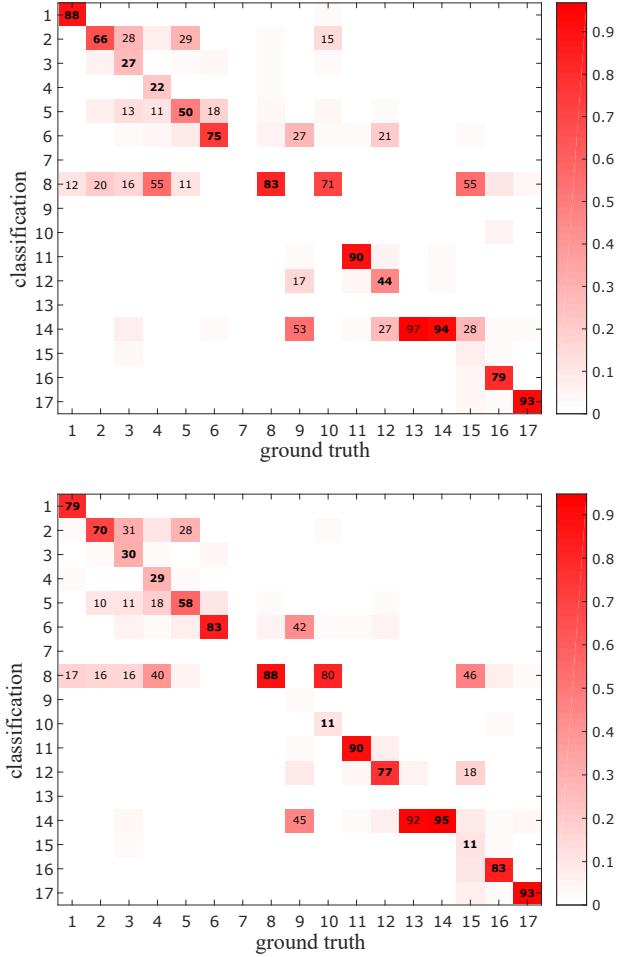


Fig. 11. Distributions of producer accuracies of the DFC dataset from the baseline framework (above) and the proposed framework (below). Values in the figure show the percentage of samples labeled as A in the ground truth data, which were classified as B in the classification maps, where A, B = 1, 2, ..., 17. Only the percentages above 10 were shown.

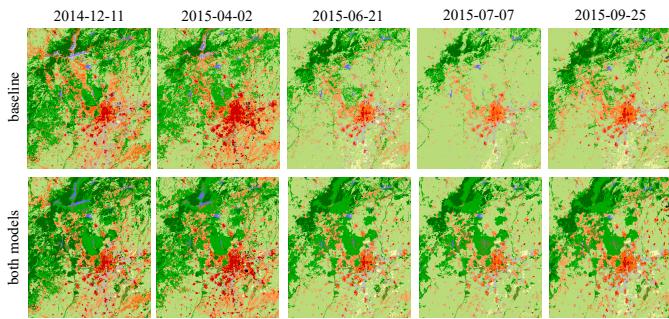


Fig. 12. Classification maps at each acquisition time in Madrid.

in terms of OA and kappa.

First, the OA and kappa from the proposed framework were 71.97% and 0.66, which outperformed the baseline accuracy by 2.22% and 3%, respectively. Given the fact that the training samples were from grss\_dfc\_2017 [39], the classification results demonstrated that our proposed framework was not only effective on four test cities from grss\_dfc\_2017 [39] but also showed satisfactory results on the newly selected

test cities. Thus, the proposed framework shows advantages of transferring knowledge from the training samples in one dataset to the test samples in the other dataset based on the same classification scheme (i.e., LCZs) and data sources (i.e., satellite images and OSM data).

Second, the building fusion model improved the accuracies more significantly than the landuse fusion model. Compared with the baseline framework, the OA using the building fusion model increased by 2.12% and 1.43% in Munich and New York, respectively; the OA using the landuse fusion model slightly increased in Munich (0.57%) and slightly decreased (0.26%) in New York. Similar phenomenon also happened in the experimental results of the DFC dataset, where OA of Chicago slightly increased using the landuse fusion model and slightly decreased using the building fusion model. The joint use of both models, however, always increases the accuracies significantly and is consistently better than using a single model.

In spite of the improvement of accuracies through the proposed framework, classification accuracies of New York were notably lower than those in Munich. The possible reason could be that our proposed fusion models are based on statistical knowledge (the probability of a value in landuse/buildings layers belongs to an LCZ's label), which needs enough training samples. However, our training samples were quite limited due to the availability of a few training cities and the lack of OSM data. Moreover, OSM data may contain errors because they are open source and could be recorded by any volunteers. Another reason could be that the weighting process used in the fusion models is simple (i.e., linear) so that it may not be satisfied when dealing with complicated cases (i.e. highly non-linear).

Fig. 15 shows the improvement of the distributions of PAs for each label. First, PAs increased on most labels, especially on label 3 (compact low-rise, 19.12%), label 4 (open high-rise, 50.58%), and label 12 (scattered trees, 29.01%). Besides, PAs of label 5 (open mid-rise, 11.94%) and label 13 (bush or scrub, 9.76%) also increased significantly. Several samples of label 1 (compact high-rise) were misclassified as label 2 (compact mid-rise), resulting in the drop of PA on label 1. This indicated that the similarity of the building density is still a challenge to acquire satisfactory separation of all built types. Beside label 1, samples of label 9 (sparsely built) were easily misclassified to label 6 (open low-rise) and label 12 (scattered trees), likely due to the classification scale as we have mentioned in the experimental analysis of the DFC dataset.

Similar to the classification results of the DFC dataset, accuracies of some labels are still not satisfactory. The separation of different built types is still a big challenge mainly because of the complexity and diversity of urban structures and classification scale. Detailed analysis has been mentioned in IV. B. 1).

### C. Framework Transferability

Spectral information plays an important role in LCZ classification when using optical satellite images, but it is quite sensitive to acquisition conditions, such as time, angle, atmospheric conditions, etc. This sensitivity decreases the robustness of

TABLE VI

CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK (ADDITIONAL TEST DATASET)

City	Baseline	Landuse Fusion Model	Building Fusion Model	Both Models
Munich	OA=89.30 K=0.86	OA=89.87 K=0.87	OA=91.42 K=0.89	OA=92.51 K=0.90
	OA=62.26 K=0.54	OA=62.00 K=0.54	OA=63.69 K=0.56	OA=64.11 K=0.57
All	OA=69.75 K=0.63	OA=69.71 K=0.63	OA=71.37 K=0.65	OA=71.97 K=0.66

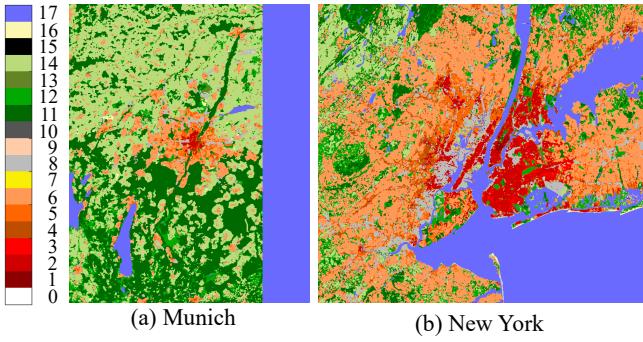


Fig. 13. Classification maps from proposed framework (additional test dataset).

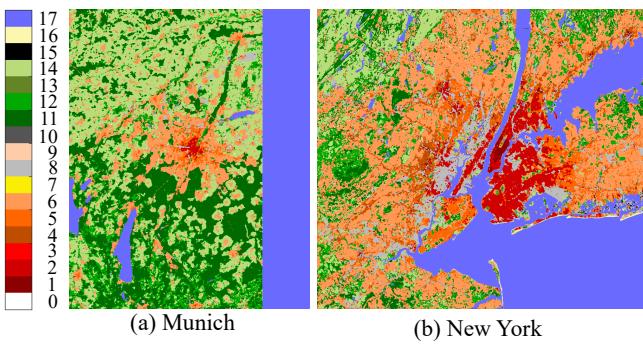
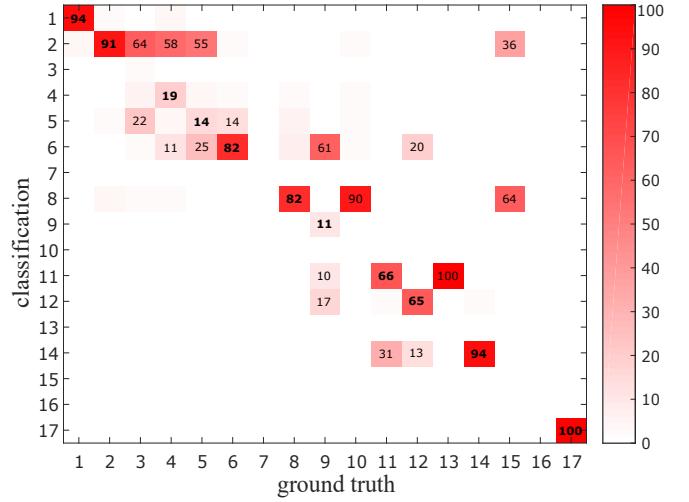


Fig. 14. Classification maps from baseline framework (additional test dataset).

the classification performance, especially when considering transferability among multi-temporal images in several study areas.

1) *Data Fusion Contest Dataset*: Tables VII, VIII, IX, X and Fig. 16 compare the classification accuracies between the proposed framework and the baseline framework using four test cities at different acquisition times. Results indicated that classification accuracies changed significantly among different acquisition times when applying the baseline framework to a test city. The difference in classification accuracies is still quite large, even if the acquisition times of satellite images are quite close (e.g., Amsterdam on March 12, 2015 and April 20, 2015), which indicates that this sensitivity is not only aroused by ground change but also due to other acquisition conditions.

After applying both fusion models to test cities, this sensitivity among different acquisition times of a test city was

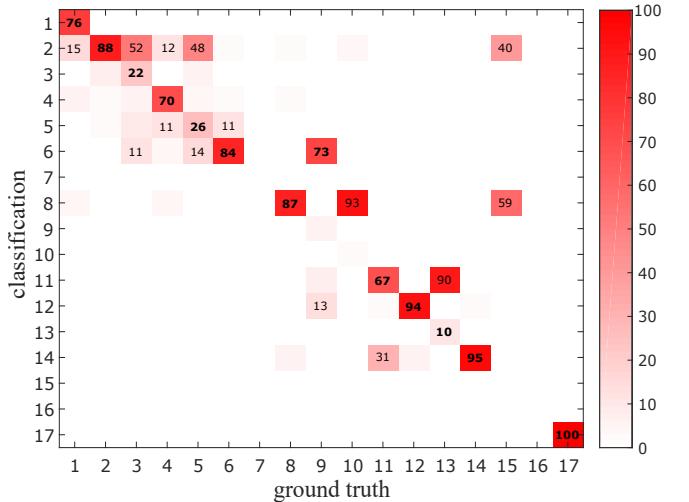


Fig. 15. Distributions of the producer accuracies of the additional test dataset from the baseline framework (above) and the proposed framework (below). Values in the figure show the percentage of samples labeled as A in the ground truth data, which were classified as B in the classification maps, where A, B = 1,2,...,17. Only the percentages above 10 were shown.

reduced significantly. Meanwhile, classification accuracies improved at all acquisition times. The reason for this significant improvement is because OSM data offer positive contributions after using the proposed fusion models. Compared to satellite images, OSM data are not sensitive to the acquisition conditions of satellite imagery since humans can provide more advanced knowledge on recognition of the ground. This

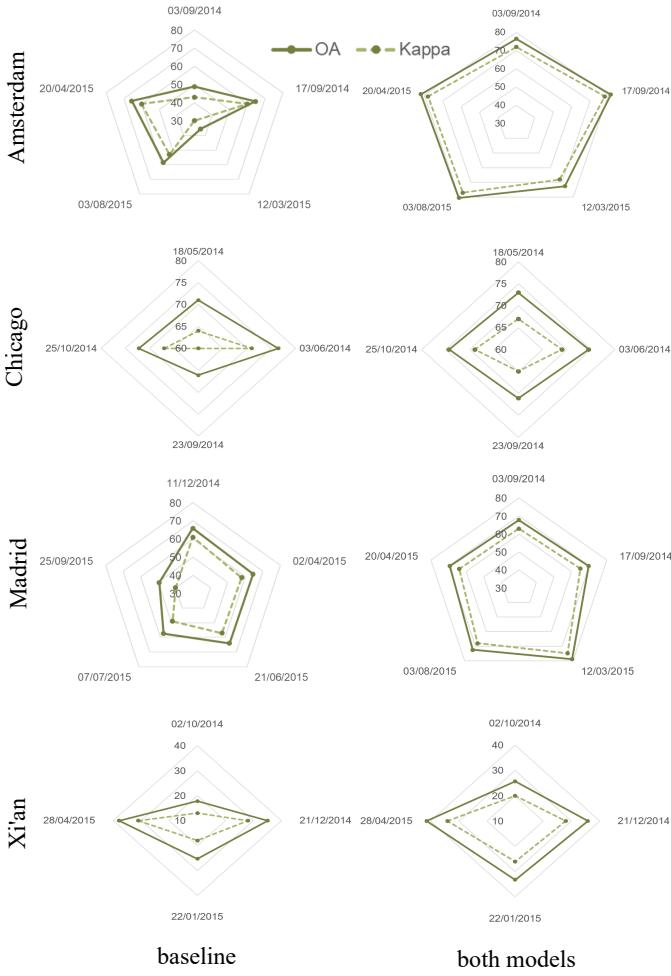


Fig. 16. Comparison of classification accuracies at different acquisition times (DFC dataset)

knowledge effectively tunes the classification results computed from satellite images and stabilizes the classification results aroused by spectral diversity.

Furthermore, the increased transferability of the proposed framework allows higher classification performance with less temporal information. Table XI compares the classification accuracies after applying frameworks to the multi-temporal Landsat-8 group<sup>3</sup> and single-temporal Sentinel-2 group<sup>4</sup> of test cities. When applying frameworks to multi-temporal data, frameworks conduct majority voting among different acquisition times of each test city.

After applying the baseline framework to all test cities, the OA difference between the Landsat-8 and Sentinel-2 groups is about 5.61%. However, this difference drops to around 0.84% after applying the proposed framework to all test cities. The city of Xi'an contributes greatly here as the OA of Xi'an after applying the Sentinel-2 group is much higher than the OA after applying the Landsat-8 group, which is a quite interesting phenomenon for further studies. Moreover, we compare the

<sup>3</sup>Multi-temporal Landsat-8 group contains multi-temporal Landsat-8 images and OSM data.

<sup>4</sup>Single-temporal Sentinel-2 group contains single-temporal Sentinel-2 images and OSM data.

OA difference in three other cities in order to remove the impact of Xi'an. After applying the baseline framework to Amsterdam, Chicago, and Xi'an, the OA difference of the three test cities, between the Landsat-8 and Sentinel-2 groups, is about 9.1%. However, this difference drops to around 3.94% after applying the proposed framework to these three test cities. These results indicate that the proposed framework could significantly improve classification performance and acquire more trustworthy classification results with less temporal information. This advantage is quite useful for urban classification when ground change should be prevented. Multi-temporal data may boost classification performance by taking temporal-spectral variability into consideration [31]. Meanwhile, when applying frameworks to multi-temporal data, the ground truth may not correspond to all data due to ground change along different acquisition times, which becomes a considerable problem. The proposed framework significantly shrinks the gap between using multi-temporal and single-temporal data. Therefore, single-temporal data could also achieve satisfactory classification performance if multi-temporal data are not available or they contain many ground changes.

Although the proposed framework could reduce the gap between OAs from the Landsat-8 group and Sentinel-2 group, classification performance is still stronger when using multi-temporal, rather than single-temporal data for the following reasons. First, multi-temporal data provides more opportunities to record the ground reflectance in several data acquisition times so that it improves the transferability between cities by increasing the spectral diversity of training and test cities. For example, the spectral reflectance of the same ground object may change occasionally due to various data acquisition conditions, thus single temporal images may not fully represent the spectral information of ground objects. Second, multi-temporal images could alleviate the cloud impact. The areas covered by clouds contain limited ground information, but it can be assumed that the clouds appear in different areas at different acquisition times. The majority voting among several classification maps generated from multi-temporal images could mostly remove the cloud impact. For example, the OA from Sentinel-2 in Chicago is significantly lower than the OA from Landsat-8, likely because the Sentinel-2 images in Chicago have high cloud coverage.

2) *Addition Test Dataset:* Tables XII, XIII, and Fig. 17 compare the classification accuracies of the proposed framework and the baseline framework using two test cities at different acquisition times. Similar to the results on the DFC dataset, classification accuracies of the baseline framework on the additional test cities were sensitive to the acquisition times of the satellite images. For example, the kappa of Munich in April and July 2017 was 0.84 and it decreased to 0.78 in October 2017. Kappa values of New York in June and October 2017 were 0.56 and 0.54, respectively, and the kappa value of New York in April 2018 dropped to 0.49.

After applying the proposed framework to test cities, the classification results were less sensitive to different acquisition times. Meanwhile, classification accuracies improved at all acquisition times. For example, kappa values of Munich at three acquisition times ranged from 0.89 to 0.90, and the kappa

TABLE VII  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN AMSTERDAM

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Landsat - 8	2014-03-09	OA=48.90 K=0.43	OA=67.19 K=0.62	OA=60.43 K=0.55	OA=76.41 K=0.72
	2014-09-17	OA=64.57 K=0.60	OA=74.95 K=0.71	OA=69.50 K=0.65	OA=81.31 K=0.78
	2015-03-12	OA=35.60 K=0.30	OA=61.26 K=0.56	OA=47.62 K=0.41	OA=72.59 K=0.68
	2015-08-03	OA=58.51 K=0.53	OA=71.81 K=0.67	OA=66.08 K=0.61	OA=80.46 K=0.77
	2015-04-20	OA=65.52 K=0.60	OA=77.21 K=0.73	OA=68.21 K=0.63	OA=81.88 K=0.78
Sentinel - 2	2016-09-08	OA=59.76 K=0.54	OA=70.91 K=0.66	OA=70.34 K=0.66	OA=81.98 K=0.78

TABLE VIII  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN CHICAGO

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Landsat - 8	2014-05-18	OA=71.01 K=0.64	OA=71.07 K=0.65	OA=70.85 K=0.64	OA=72.97 K=0.67
	2014-06-03	OA=76.46 K=0.71	OA=72.31 K=0.66	OA=75.11 K=0.70	OA=74.56 K=0.69
	2014-09-23	OA=66.09 K=0.6	OA=66.78 K=0.66	OA=70.10 K=0.64	OA=71.11 K=0.65
	2014-10-25	OA=72.21 K=0.67	OA=69.42 K=0.63	OA=73.67 K=0.68	OA=74.31 K=0.69
	2016-10-13	OA=55.25 K=0.49	OA=55.61 K=0.49	OA=69.57 K=0.64	OA=69.63 K=0.64

TABLE IX  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN MADRID

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Landsat - 8	2014-12-11	OA=65.92 K=0.61	OA=67.80 K=0.63	OA=65.14 K=0.60	OA=67.80 K=0.63
	2015-04-02	OA=64.44 K=0.58	OA=68.85 K=0.64	OA=62.69 K=0.57	OA=69.70 K=0.65
	2015-06-21	OA=63.94 K=0.57	OA=78.07 K=0.74	OA=64.50 K=0.58	OA=78.94 K=0.75
	2015-07-07	OA=57.35 K=0.49	OA=69.73 K=0.64	OA=57.61 K=0.49	OA=72.52 K=0.68
	2015-09-25	OA=49.38 K=0.40	OA=67.48 K=0.62	OA=51.23 K=0.42	OA=69.23 K=0.64
Sentinel - 2	2016-10-11	OA=68.96 K=0.63	OA=73.72 K=0.69	OA=67.21 K=0.61	OA=74.92 K=0.71

values of New York ranged from 0.53 to 0.59. Therefore, it can be concluded that the proposed framework also demonstrated satisfactory transferability on the additional test dataset. The detailed analysis of the framework transferability can be found in IV. C. 1).

#### D. Building Confidence Masks

Building confidence masks remove most incomplete building areas and provide much cleaner building density features for generating the building weight matrix. In this section, we generate building weight matrices using different building confidence masks with various approaches and thresholds (Fig. 19). Fig. 19 (a) to (k) present the matrices after applying the

TABLE X  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN XI'AN

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Landsat - 8	2014-10-02	OA=17.78 K=0.13	OA=25.15 K=0.19	OA=16.92 K=0.12	OA=25.72 K=0.20
	2014-12-21	OA=35.05 K=0.28	OA=35.41 K=0.28	OA=31.26 K=0.24	OA=35.81 K=0.28
	2015-01-22	OA=25.28 K=0.18	OA=32.48 K=0.25	OA=23.35 K=0.16	OA=33.22 K=0.26
	2015-04-28	OA=37.88 K=0.31	OA=41.27 K=0.34	OA=36.49 K=0.29	OA=41.35 K=0.34
	Sentinel - 2	2016-08-27	OA=53.33 K=0.44	OA=59.31 K=0.52	OA=58.00 K=0.45

TABLE XI  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN LANDSAT-8 AND SENTINEL-2

Satellite	City	Baseline	Landuse Model	Building Model	Both Models
Landsat - 8	Amsterdam	OA=59.74 K=0.55	OA=72.34 K=0.68	OA=68.08 K=0.63	OA=81.12 K=0.77
	Chicago	OA=78.15 K=0.73	OA=76.43 K=0.71	OA=78.80 K=0.74	OA=78.56 K=0.74
	Madrid	OA=73.37 K=0.68	OA=78.18 K=0.74	OA=72.71 K=0.67	OA=78.67 K=0.75
	Xi'an	OA=37.16 K=0.31	OA=37.74 K=0.31	OA=31.61 K=0.25	OA=38.48 K=0.31
	All	OA=67.08 K=0.62	OA=70.44 K=0.66	OA=67.17 K=0.62	OA=72.51 K=0.68
Sentinel - 2	Amsterdam	OA=59.76 K=0.54	OA=70.91 K=0.66	OA=70.34 K=0.66	OA=81.98 K=0.78
	Chicago	OA=55.25 K=0.49	OA=55.61 K=0.49	OA=69.57 K=0.64	OA=69.63 K=0.64
	Madrid	OA=68.96 K=0.63	OA=73.72 K=0.69	OA=67.21 K=0.61	OA=74.92 K=0.71
	Xi'an	OA=53.33 K=0.44	OA=59.31 K=0.52	OA=53.67 K=0.45	OA=58.00 K=0.50
	All	OA=61.47 K=0.56	OA=66.05 K=0.61	OA=66.10 K=0.61	OA=71.67 K=0.67

TABLE XII  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN MUNICH

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Sentinel-2	2017-04-24	OA=87.71 K=0.84	OA=89.91 K=0.87	OA=88.00 K=0.85	OA=92.29 K=0.90
	2017-07-18	OA=87.75 K=0.84	OA=89.02 K=0.86	OA=89.57 K=0.87	OA=91.24 K=0.89
	2017-10-16	OA=82.24 K=0.78	OA=90.05 K=0.87	OA=85.50 K=0.82	OA=91.79 K=0.90

proposed approach to generating building confidence masks (the thresholds of the building surface fraction range from 0% to 100% with the step of 10%). Fig. 19 (l) indicates the matrix after using the rule of the building surface fraction (Table III) to generate building confidence masks. Fig. 19 (m) demonstrates the matrix after using all-pass masks. All ma-

trices were computed by using training samples as computing the matrix (l) in Fig. 19 required ground truth data. The matrix (l) in Fig. 19 could be regarded as quasi-truth. It is assumed that an effective approach to generating building confidence masks should have a high correlation with the matrix (l) in Fig. 19 and a low correlation with the matrix (m) in Fig. 19.

TABLE XIII  
CLASSIFICATION ACCURACIES COMPARISON BETWEEN THE PROPOSED FRAMEWORK AND BASELINE FRAMEWORK AT EACH ACQUISITION TIME IN NEW YORK

Satellite	Date	Baseline	Landuse Model	Building Model	Both Models
Sentinel-2	2017-06-12	OA=63.78 K=0.56	OA=63.38 K=0.56	OA=67.86 K=0.61	OA=66.25 K=0.59
	2017-10-20	OA=61.45 K=0.54	OA=63.83 K=0.57	OA=63.67 K=0.56	OA=64.05 K=0.57
	2018-04-23	OA=57.83 K=0.49	OA=59.59 K=0.51	OA=60.93 K=0.53	OA=61.60 K=0.53

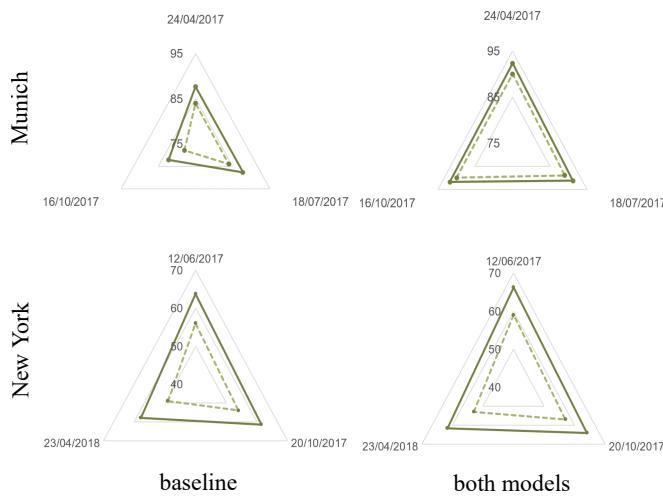


Fig. 17. Comparison of classification accuracies at different acquisition times (additional test dataset)

The correlation between the matrices generated from different building confidence masks are investigated in Fig. 20. Fig. 20 (a) presents the correlation between the matrices of Fig. 19 (a) – (k) and the matrix of Fig. 19 (l) and (b) illustrates the correlation between the matrices of Fig. 19 (a) – (k) and the matrix of Fig. 19 (m).

The results indicate that, firstly, building confidence masks are not sensitive to the threshold of a building surface fraction between 10 and 70. Second, the correlation significantly increased from a threshold of 0 to 10 and then remained stable between 10 and 70, indicating that a threshold of 10 is acceptable for this paper. Third, the correlation became unsatisfied after a threshold of 70 because the high threshold masked out many areas that should have been retained. Fourth, these two correlation graphs not only indicate that the generation of building confidence masks is not sensitive to the threshold of a building surface fraction but also provide a way to automatically select the threshold. For example, a threshold of 50 should be selected to retain a larger correlation in Fig. 20 (a) and a lower correlation in Fig. 20 (b). We set the threshold as 10 for the following two reasons. First, this parameter is not sensitive. Second, sparsely built areas can be retained, because 10 % is the lowest boundary of the building surface fraction of that label.

Although this approach is effective from a statistical point of view, several open issues remain. Certain empirical parameters

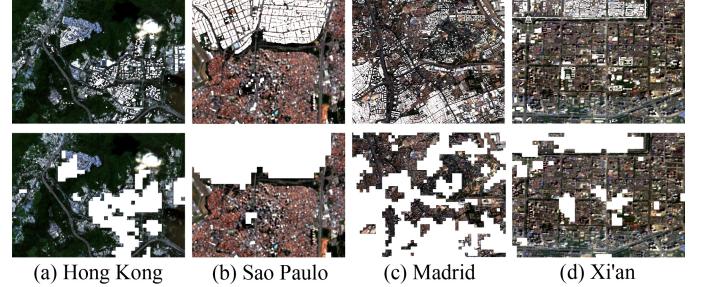


Fig. 18. Examples of building confidence masks (upper: a buildings layer overlaid with Sentinel-2 images; below: a building confidence mask overlaid with Sentinel-2 images).

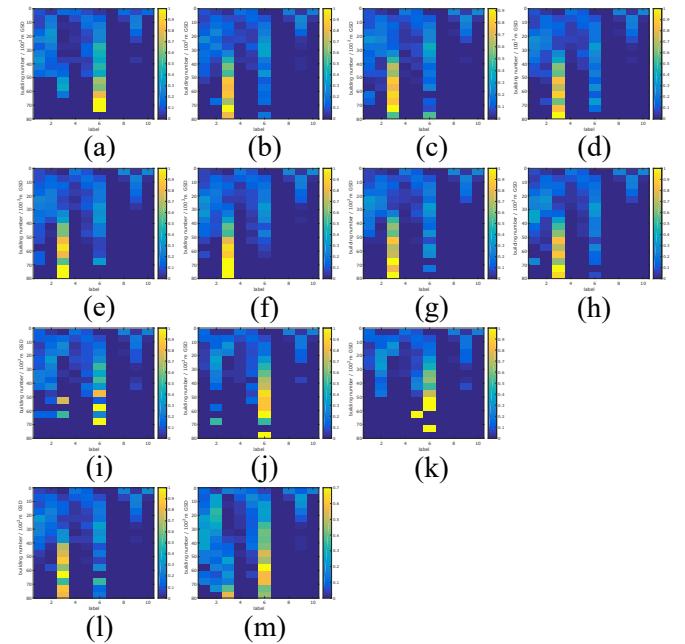


Fig. 19. A comparison of building densities – label distributions among different approaches to generating building confidence masks under various thresholds of building surface fraction. (a)-(k) proposed approach using the building surface fraction thresholds from 0% to 100% with a step of 10%; (l) building surface fraction rule; (m) all-pass masks.

are included, such as the searching radius, threshold of the building surface fraction, and threshold of generating masks. Since very high-resolution images are not available in the dataset [39], this approach still initiates a novel idea about building data validation by considering the relations among OSM data.

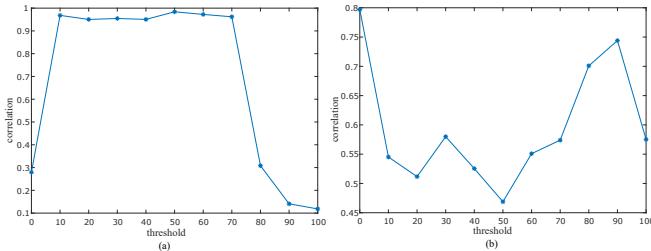


Fig. 20. Correlation between different building density - label distributions. (a) between Fig. 19 (a) - (k) and Fig. 20 (l); (b) between Fig. 19 (a) - (k) and Fig. 20 (m).

## V. DISCUSSION

The proposed framework with two fusion models acquires higher classification accuracies while additionally achieving more stable and generalized results on a worldwide scale. First, feature stacking is the most commonly used and fastest implemented approach of feature fusion. It stacks features along the feature dimension and with the intent that the classifier(s) will find the satisfactory hyperplane, but this is not always the case. When features are heterogeneous, such as having different noise sources, it will add many outliers in the feature space, which significantly impacts hyperplane splits. This paper has resolved this heterogeneity problem by embedding the relation among multi-source data in the proposed fusion models. This presents a novel idea to fuse different data sources by considering both features and their deep relations.

Second, it is of great interest to understand how the data sources achieve optimal collaboration. OSM data containing human intelligence should have more advanced and stable knowledge of the ground objects, compared with satellite images, but they contain incomplete and incorrect recordings. Satellite images recorded by sensors typically have more objective and abundant information from the ground; however, they are quite sensitive to acquisition conditions. In order to stabilize the classification performance, the proposed framework firstly creates an initial classification result through satellite images and then tunes the initial result by using OSM data.

The proposed framework still contains several open problems. First, the proposed fusion models use a simple 2D probability distribution to map the relation between OSM and satellite features, which may be not accurate. Therefore, more complicated models that can embed deeper relations, such as association rule learning [44], can be derived. Second, to avoid a negative impact, we have resolved the problems of the data incompleteness of OSM, but we omitted the problem of incorrect recordings from OSM data. Third, certain empirical parameters are needed to generate building confidence masks. Fourth, because of multi-source and multi-temporal data, the acquisition times of all data sources are not consistent, which could lead to many problems in analyzing the data without discrimination. Fourth, the model transferability between cities requires further research. Due to the high cost of selecting the training samples, it would be ideal to transfer the knowledge

trained from some cities to classify other cities. CCF has already successfully demonstrated how to resolve this problem, by embedding the correlation between features and labels in the projected feature space. However, this solution may not be accurate or the best method. Lastly, urban scenes contain many intermediate LCZ labels. Each label of LCZs is defined as a certain combination of several ground types with certain structure arrangements that have only a few range values, instead of quantitatively precise definitions, to describe each label [8]. This could generate many intermediate areas, which may belong to multi-labels.

## VI. CONCLUSION

This paper proposes a framework with two novel models of fusing satellite optical images and OSM data for the classification of LCZs. The contributions of extracted features have been investigated, and it has been discovered that OSM features possess trivial or even negative contributions to the classification performance through the use of feature stacking. Accordingly, we proposed a new framework that embeds the multi-source data and their relations by considering the data heterogeneity of optical images and OSM data. The proposed framework achieves better classification performance than state-of-the-art frameworks. Furthermore, its increased robustness shows promise in generalizing the framework for a worldwide scale with less temporal information use. In addition, this paper introduces the novel idea of detecting incomplete data areas in buildings layers by considering their relation with landuse layers. This approach offers much assistance in building incompleteness detection when very high-resolution images are not available.

This paper contains certain open issues that are of interest for further investigation. First, there is the prospect of forming more sophisticated fusion models that embed additional complex rules between heterogeneous data. Specifically, it can be an excellent future work to extend this framework to fuse other data sources, such as SAR images. Second, solving the problem of incomplete and inaccurate recordings in OSM data remains a challenging topic. Third, there is a need for the investigation of knowledge transferability among different study regions in order to generate classification maps on a global scale.

## ACKNOWLEDGMENT

The authors would like to thank the WUDAPT (<http://www.wudapt.org/>) and GeoWIKI (<http://geowiki.org/>) initiatives for providing the data packages used in this study, the DASE benchmarking platform (<http://dase.ticinumaerospace.com/>), and the IEEE GRSS Image Analysis and Data Fusion Technical Committee. Landsat 8 data are available from the U.S. Geological Survey (<https://www.usgs.gov/>). OpenStreetMap Data OpenStreetMap contributors, available under the Open Database Licence <http://www.openstreetmap.org/copyright>. Original Copernicus Sentinel Data 2016 are available from the European Space Agency (<https://sentinel.esa.int>).

## REFERENCES

- [1] W. P. Lowry, "Empirical estimation of urban effects on climate: a problem analysis," *Journal of applied meteorology*, vol. 16, no. 2, pp. 129–135, 1977.
- [2] S. D. Brunn, J. Williams, and D. J. Zeigler, *Cities of the world: world regional urban development*. Rowman & Littlefield, 2003.
- [3] G. C.-S. Lin, "Changing theoretical perspectives on urbanisation in asian developing countries," *Third World Planning Review*, vol. 16, no. 1, p. 1, 1994.
- [4] B. Bechtel, P. J. Alexander, J. Böhner, J. Ching, O. Conrad, J. Feddema, G. Mills, L. See, and I. Stewart, "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS International Journal of Geo-Information*, vol. 4, no. 1, pp. 199–219, 2015.
- [5] M. Jin, R. E. Dickinson, and D. Zhang, "The footprint of urban areas on global climate as characterized by modis," *Journal of climate*, vol. 18, no. 10, pp. 1551–1565, 2005.
- [6] T. Esch, M. Marconcini, A. Felbier, A. Roth, W. Heldens, M. Huber, M. Schwinger, H. Taubenböck, A. Müller, and S. Dech, "Urban footprint processorfully automated processing chain generating settlement masks from global data of the tandem-x mission," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1617–1621, 2013.
- [7] B. Bechtel and C. Daneke, "Classification of local climate zones based on multiple earth observation data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1191–1202, 2012.
- [8] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bulletin of the American Meteorological Society*, vol. 93, no. 12, pp. 1879–1900, 2012.
- [9] I. D. Stewart, T. R. Oke, and E. S. Krayenhoff, "Evaluation of the local climate zone scheme using temperature observations and model simulations," *International Journal of Climatology*, vol. 34, no. 4, pp. 1062–1080, 2014.
- [10] OpenStreetMap contributors, "Planet dump retrieved from <https://planet.osm.org/>," <https://www.openstreetmap.org>, 2017.
- [11] B. A. Johnson, K. Izuka, M. A. Bragais, I. Endo, and D. B. Magcale-Macandog, "Employing crowdsourced geographic data and multi-temporal/multi-sensor satellite imagery to monitor land cover change: A case study in an urbanizing region of the philippines," *Computers, Environment and Urban Systems*, vol. 64, pp. 184–193, 2017.
- [12] N. Audebert, B. Le Saux, and S. Lefèvre, "Joint learning from earth observation and OpenStreetMap data to get faster better semantic maps," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1552–1560, 2017.
- [13] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
- [14] 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 International Journal of Geo-Information*, vol. 7, no. 9, p. 379, 2018.
- [15] C. Qiu, M. Schmitt, L. Mou, P. Ghamisi, and X. Zhu, "Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets," *Remote Sensing*, vol. 10, no. 10, p. 1572, 2018.
- [16] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing*, vol. 28, no. 5, pp. 823–870, 2007.
- [17] D. Tuia, G. Moser, B. Le Saux, B. Bechtel, and L. See, "2017 IEEE GRSS data fusion contest: Open data for global multimodal land use classification [technical committees]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 70–73, 2017.
- [18] J. R. Anderson, *A land use and land cover classification system for use with remote sensor data*. US Government Printing Office, 1976, vol. 964.
- [19] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: an applied review," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, 2018.
- [20] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 1, pp. 8–32, 2017.
- [21] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, Dec 2017.
- [22] C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of remote sensing*, vol. 23, no. 4, pp. 725–749, 2002.
- [23] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sánchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012.
- [24] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.
- [25] T. Rainforth and F. Wood, "Canonical correlation forests," *arXiv preprint arXiv:1507.05444*, 2015.
- [26] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.
- [27] G. Mills, J. Ching, L. See, B. Bechtel, and M. Foley, "An introduction to the WUDAPT project," in *Proceedings of the 9th International Conference on Urban Climate, Toulouse, France*, 2015, pp. 20–24.
- [28] N. Kaloustian, M. Tamminga, and B. Bechtel, "Local climate zones and annual surface thermal response in a mediterranean city," in *2017 Joint Urban Remote Sensing Event (JURSE)*, March 2017, pp. 1–4.
- [29] B. Bechtel, O. Conrad, M. Tamminga, M.-L. Verdonck, F. Van Coillie, D. Tuia, M. Demuzere, L. See, P. Lopes, C. C. Fonte et al., "Beyond the urban mask," in *Urban Remote Sensing Event (JURSE), 2017 Joint IEEE*, 2017, pp. 1–4.
- [30] P. Lopes, C. Fonte, L. See, and B. Bechtel, "Using OpenStreetMap data to assist in the creation of LCZ maps," in *Urban Remote Sensing Event (JURSE), 2017 Joint IEEE*, 2017, pp. 1–4.
- [31] N. Yokoya, P. Ghamisi, and J. Xia, "Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 1197–1200.
- [32] Y. Xu, F. Ma, D. Meng, C. Ren, and Y. Leung, "A co-training approach to the classification of local climate zones with multi-source data," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 1209–1212.
- [33] S. Sukhanov, I. Tankoyeu, J. Louradour, R. Heremans, D. Trofimova, and C. Debes, "Multilevel ensembling for local climate zones classification," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 1201–1204.
- [34] C. S. dos Anjos, M. G. Lacerda, L. do Livramento Andrade, and R. N. Salles, "Classification of urban environments using feature extraction and random forest," in *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, 2017, pp. 1205–1208.
- [35] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 41–57, 2016.
- [36] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.
- [37] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. Le Saux, G. Moser, and D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–15, 2018.
- [38] L. Gomez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, "Multimodal classification of remote sensing images: A review and future directions," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.
- [39] 2017 IEEE GRSS data fusion contest. Online: <http://www.grss-ieee.org/community/technical-committees/data-fusion> [Accessed: 06.10.2017]. [Online]. Available: <http://www.grss-ieee.org/community/technical-committees/data-fusion>
- [40] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. Le Saux, G. Moser, and D. Tuia. Additional material to jstars-2017-1043: Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest. [Accessed: 31.01.2018]. [Online]. Available: <http://www.grss-ieee.org/wp-content/uploads/2018/01/ieee-grss-iadf-dfc2017-jstars-suppl-material.pdf>
- [41] F. Ramm, "OpenStreetMap data in layered GIS format," *Version 0.6*, vol. 7, 2014.
- [42] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of sentinel-2 images," *Remote sensing of environment*, vol. 187, pp. 241–252, 2016.
- [43] X. Zhang, J. Cui, W. Wang, and C. Lin, "A study for texture feature extraction of high-resolution satellite images based on a direction

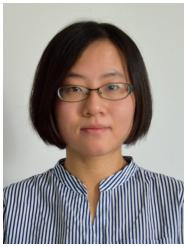
- measure and gray level co-occurrence matrix fusion algorithm,” *Sensors*, vol. 17, no. 7, p. 1474, 2017.
- [44] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” in *AcM sigmod record*, vol. 22, no. 2. ACM, 1993, pp. 207–216.



**Xiao Xiang Zhu** (S’10–M’12–SM’14) received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her Habilitation in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively.

She is currently the Professor for Signal Processing in Earth Observation ([www.sipeo.bgu.tum.de](http://www.sipeo.bgu.tum.de)) at Technical University of Munich (TUM) and German Aerospace Center (DLR); the head of the department “EO Data Science” at DLR’s Earth Observation Center; and the head of the Helmholtz Young Investigator Group “SiPEO” at DLR and TUM. Since 2019, Zhu is co-coordinating the Munich Data Science Research School ([www.mu-ds.de](http://www.mu-ds.de)). She is also leading the Helmholtz Artificial Intelligence Cooperation Unit (HAICU) – Research Field “Aeronautics, Space and Transport”. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, the University of Tokyo, Tokyo, Japan and University of California, Los Angeles, United States in 2009, 2014, 2015 and 2016, respectively. Her main research interests are remote sensing and Earth observation, signal processing, machine learning and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing.



**Guichen Zhang** received the B.Sc. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the M.Sc. degree in earth oriented space science and technology with the Technical University of Munich, Munich, Germany, in 2018. Since 2014, she has been a master student in photogrammetry and remote sensing with Wuhan University, Wuhan, China. Since 2018, she has been pursuing the Ph.D. degree with the German Aerospace Center, Wessling, Germany. Her research interests include image analysis and data fusion in the application of remote sensing.



**Pedram Ghamisi** (S12–M15–SM12) received the B.Sc. degree in civil (survey) engineering from the Tehran South Campus of Azad University, Tehran, Iran, in 2008, the M.Sc. degree (Hons.) in remote sensing with the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the Ph.D. degree in electrical and computer engineering with the University of Iceland, Reykjavik, Iceland, in 2015.

In 2013 and 2014, he was with the School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, QLD, Australia. In 2015, he was a Post-Doctoral Research Fellow with the Technical University of Munich (TUM), Munich, Germany, and Heidelberg University, Heidelberg, Germany. From 2015 to 2018, he was a Research Scientist with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Since 2018, he has been working as the Head of the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR). He is also the CTO and the Co-Founder of VasoGnosis Inc., Milwaukee, WI, USA, where he is involved in the development of advanced diagnostic and analysis tools for brain diseases using cloud computing and deep learning algorithms. His research interests include interdisciplinary research on remote sensing and machine (deep) learning, image and signal processing, and multisensor data fusion.

Dr. Ghamisi was a recipient of the Best Researcher Award for M.Sc. students at the K. N. Toosi University of Technology in the academic year of 2010–2011, the IEEE Mikio Takagi Prize for winning the Student Paper Competition at IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2013, the Talented International Researcher by Irans National Elites Foundation in 2016, the first prize of the data fusion contest organized by the Image Analysis and Data Fusion Technical Committee (IADF) of IEEE-GRSS in 2017, the Best Reviewer Prize of IEEE Geoscience and Remote Sensing Letters (GRSL) in 2017, the Alexander von Humboldt Fellowship from the Technical University of Munich, and the High Potential Program Award from HZDR. He serves as an Associate Editor for Remote Sensing and IEEE GRSL.