Exploring How Street-Level Images Help Enhance Remote-Sensing-Based Local Climate Zone Mapping

Cai Liao , Rui Cao , *Member, IEEE*, Qi-Li Gao , Jinzhou Cao , and Nianxue Luo

Abstract—The local climate zone (LCZ) classification scheme is effective for climatic studies, and thus, timely and accurate LCZ mapping becomes critical for scientific climate research. Remote sensing images can efficiently capture the information of large-scale landscapes overhead, while street-level images can supplement the ground-level information, thus helping improve the LCZ mapping. Previous study has proven the usefulness of street-level images in enhancing LCZ mapping results; however, how they help to improve the results still remains unexplored. To unveil the underlying mechanism and fill the gap, in this study, the feature importance analysis is performed on classification experiments using different data sources to reveal the contributions of different components, while feature correlation analysis is adopted to find the relationship between street view images and key LCZ indicators. The results show that fusing street view images can help improve the classification performance considerably, especially for compact urban types such as compact highrise and compact midrise. In addition, the results further show that the building and sky information embedded in the street view images contribute the most. The feature correlation analysis further demonstrates their strong correlations with key LCZ indicators, which define the LCZ scheme. The findings of the study can help us better understand how street-level images can contribute to LCZ mapping and facilitate future urban climate studies.

Index Terms—Climate change, data fusion, interpretability, local climate zone (LCZ), remote sensing, street-level images.
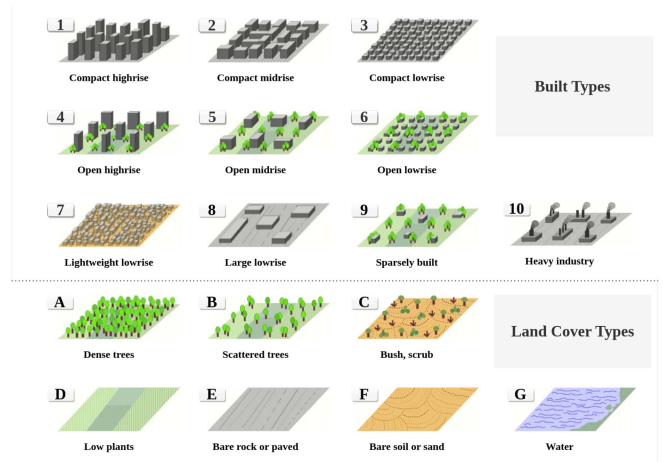
Fig. 1. LCZ classification scheme, which includes ten built classes (1–10) and seven land cover classes (A–G) [2]. (The figure is adapted from previous work [4]).

I. INTRODUCTION

THE urban heat island (UHI) phenomenon occurs when urban areas experience significantly higher temperatures

Cai Liao is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430072, China, also with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, and also with the The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen , China (e-mail: liaocai@whu.edu.cn).

Rui Cao is with the Department of Land Surveying and Geo-Informatics and Otto Poon Charitable Foundation Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong, and also with The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen , China (e-mail: rucao@polyu.edu.hk).

Qi-Li Gao is with the Centre for Advanced Spatial Analysis, University College London, WC1E 6BT London, U.K. (e-mail: qili.gao@ucl.ac.uk).

Jinzhou Cao is with the Shenzhen Technology University, Shenzhen 518118, China (e-mail: caojinzhou@sztu.edu.cn).

Nianxue Luo is with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430072, China (e-mail: nxluo@sgg.whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3301792

compared to their rural surroundings [1]. This phenomenon is closely linked to the processes of climate change and urbanization. As urban populations continue to grow, with over half of the global population residing in urban regions, the impacts of UHI are becoming a pressing concern that demands close attention.

The local climate zone (LCZ) classification system [2] provides a research framework for the study of UHI, which provides a way to categorize urban areas based on their key characteristics, including urban structure, land cover, and human activity. Specifically, the LCZ classification system consists of 17 types, including ten built types and seven land cover types, as illustrated in Fig. 1. This information can then be used to study the thermal behavior of an urban area and its impact on the local climate. Furthermore, LCZ facilitates the exchange of urban temperature observations, making it an indispensable tool for climate research in general. In addition, it is found to be useful for climate modeling, weather forecasting, urban planning, building energy, carbon emission, and many other related fields [3]. Therefore, it is important for timely and accurate LCZ mapping.

Currently, there are several ways to map LCZ; according to the major data source used, these methods can be broadly categorized into three classes, i.e., *in situ* measurement-based method [5], GIS-based method [6], [7], and remote-sensing-based [8], [9] method. Remote-sensing-based methods have become popular due to the easy accessibility to large-scale remote sensing data. However, they lack ground-level details,

especially in high-density urban areas, which is vital for LCZ mapping [10]. Street view images (SVIs), as an emerging data source, have shown the ability to capture 3-D building structure and ground-level details [11], [12] and, therefore, can serve as a valuable source of information for LCZ mapping. However, despite its potential, research on how SVIs can aid in LCZ mapping results remains limited [4], [13], [14].

To address this gap, we conduct extensive experiments to investigate how SVIs can help with LCZ mapping from different perspectives. Our study found that SVIs contain useful information, especially for building details, which are critical for LCZ recognition and mapping. The findings of this study provide valuable insights into how street-level images can contribute to LCZ mapping and offer useful guidance for future urban climate studies.

The rest of this article is organized as follows. Section II reviews the related works of LCZ mapping and SVI applications. Section III describes the study area and data. Section IV introduces the methods used. Section V presents the experiments and results analysis. Section VI discusses some important issues. Finally, Section VII concludes this article.

II. RELATED WORK

A. LCZ Mapping

Mapping LCZs has been the subject of numerous efforts, which can generally be categorized into three methods: *in situ* measurement [5], GIS-based [7], [15], and remote-sensing-based [8], [9]. *In situ* measurement requires professionals to deploy instruments in the field and collect data related to LCZ types [5]. Because this method involves complex processing and demands the involvement of personnel, it is time consuming and impractical at large scale. GIS-based methods, on the other hand, rely on complete and high-quality GIS data such as digital surface models (DSMs), road network data, and 3-D building data, which can be used to calculate LCZ-related indicators [2]. These indicators can further be leveraged to classify the land into different LCZ types according to the related classification rule sets [7], [15]. Although these methods are convenient, they tend to be difficult to implement widely because collecting complete GIS data is challenging in many cities, particularly in less developed ones.

Remote-sensing-based methods are gaining increasing attraction as remote sensing technology advances and open access data become more available. Researchers have conducted various studies using remote sensing data, and among these studies, the World Urban Database and Access Portal Tools (WUDAPT) [8] method has become a prevalent choice. The WUDAPT method acquires free Landsat remote sensing data and volunteer-contributed digitized LCZ labels to build random forest models and then performs classification prediction using the trained models [8]. Although popular, the WUDAPT method is inherently a pixel-level classification approach, which overlooks the contextual information that is critical for LCZ mapping. To address the issue, grid-based methods are proposed, which first divide the study areas into spatial grids with size of several hundreds and then classify the grids with enclosed image patches,

which include more surrounding contextual information [9], [16]. Compared with GIS-based methods, remote-sensing-based methods employ more easily available satellite images, making it easier to update and manage LCZ maps.

B. SVI Applications

SVIs are collected with horizontal view points that are similar to human beings and can capture abundant ground-level information, which usually cannot be captured from the top-down view by satellite imagery or aerial imagery. Therefore, they are considered as a valuable source to complement remote sensing imagery (RSI). In addition, with the growing accessibility of SVIs and the image understanding and computer vision techniques, SVIs are widely used for a wide variety of applications [11], such as urban land use classification [17], [18], building instance classification [19], building height estimation [12], house price prediction [20], crop type identification [21], mobility pattern prediction [22], and health-related research [23]. Previous research has also shown that street view images are useful for LCZ classification in image level [13], [14], and they are also demonstrated to be complementary to RSI in LCZ mapping [4].

C. Summary

Owing to the capture view limitations, remote sensing images cannot capture LCZ-related ground-level details, which, however, can be provided by street-level images. Previous works have demonstrated the value of SVIs for LCZ classification in sensing the urban environment [4], [13], [14] and have empirically proven the usefulness of integrating SVIs with RSI to augment the mapping performance [4]. However, how the SVIs help to achieve this improvement still remains unexplored. To bridge the gap, in this study, we present several approaches to understanding the situation from different perspectives.

III. STUDY AREA AND DATA

A. Study Area

Hong Kong is one of the most densely populated cities in the world, with over 7.5 million population residing only less than one-quarter of the total 1104-km^2 land area. It has 18 districts with a wide variety of landforms, as shown in Fig. 2. The urban morphology of Hong Kong is sophisticated, which poses challenges for automatic LCZ classification and mapping in Hong Kong.

B. Data for LCZ Mapping

1) Remote Sensing Imagery: For the RSI, the Sentinel-2 imagery is adopted in this study, which is obtained from Google Earth Engine, filtered by cloud cover percentage less than 3%. After that, ten bands are selected from the Sentinel-2 imagery following previous practice [9]. The selected bands are B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. For red, green, blue, and near infrared bands, the spatial resolution is 10 m, while others are 20 m, which are resampled to 10 m for further use.
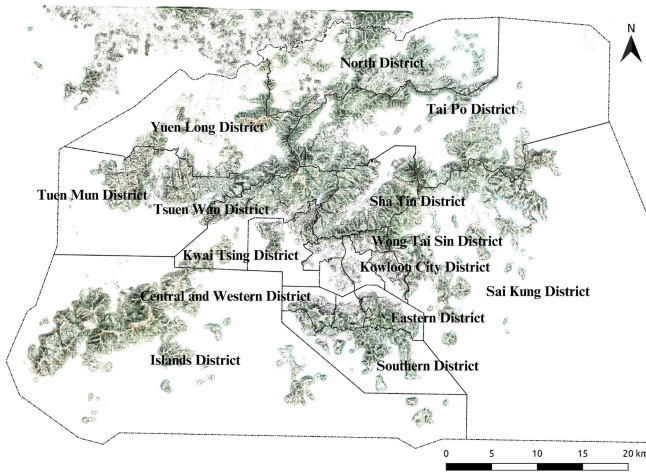
Fig. 2.   Study area of Hong Kong.

2) Street View Images: The SVIs are requested from Google Street View Static API.[1] The sample points of SVIs are generated along the road networks with a distance equal to 10 m, and the hexagon-constraint method with an inradius of 50 m is used to decrease the sample points while keeping them distributed evenly [4]. Then, with these sample points, we request the corresponding SVIs via the Google Street View API while passing the latitude and longitude of these points as parameters. To best exploit the panoramic view of SVIs, for each sample point, we request four images with field of view equal to 90° and headings of 0°, 90°, 180°, and 270°. The image size is taken as 640 × 640 pixels.

3) Auxiliary GIS Data: To perform the LCZ mapping, the Hong Kong administrative boundary is essential for defining the study area, and the mapping result is also based on the divided grids of the boundary data. We downloaded the official administrative boundary GIS data from the government geodata portal. In addition, since Google SVIs are mainly captured along road networks, we downloaded the OSM road networks of Hong Kong via the Python package OSMnx [24].

4) LCZ Label Data: In this study, LCZ label data for Hong Kong provided by the WUDAPT project [8] are used, spreading over all 18 districts in Hong Kong covering 17 LCZ categories. The data are provided in vector GIS format, where each vector element is represented as a polygonal area with an LCZ category attribute identifying the area as belonging to a certain LCZ type. To construct the classification dataset, the LCZ polygons are first rasterized and then provided as the label for the mapping units. In this study, the number of mapping units with valid LCZ labels is 2555.

C. Data for Interpretation

Although it is difficult to understand how SVIs help augment the LCZ mapping performance directly, it is helpful to relate SVIs to the key indicators of urban morphology and land cover

parameters, which are closely associated with the LCZ classification system and can serve as proxy for interpretation.

From previous studies, among all the LCZ-related indicators, surface structure and cover parameters have been commonly used for LCZ classification [25], including sky view factor (SVF), aspect ratio, building surface fraction (BSF), impervious surface fraction (ISF), pervious surface fraction (PSF), height of roughness elements, and terrain roughness class [2]. In this study, the four key commonly used indicators, i.e., SVF, BSF, ISF, and PSF, are adopted. To calculate the SVF, DSM data covering the whole Hong Kong region are collected from Esri China (Hong Kong). There are 2084 DSM tiles with spatial resolution of 0.5 m. The DSM tiles are first merged into a single tiff file and then resampled to 2 m to alleviate the computational burden. For BSF calculation, building data are collected from the Hong Kong GeoData store website. The collected building data are provided by the Hong Kong Lands Department in the form of polygons, and the geometry area of each building footprint is used to calculate the BSF value. The PSF can be derived from calculating the normalized difference vegetation index (NDVI) value using the Sentinel-2 imagery [15].

D. Study Units

In the study, there are two kinds of study units, i.e., uniform spatial grids and hexagonal grids. The former grids are used for remote sensing image crop and serve as the basic mapping units for LCZ mapping, while the latter is the sampling district for SVIs and also serve as the study unit for key LCZ indicator computation. The spatial relationships of the spatial grids and hexagons are illustrated in Fig. 3.

IV. METHODOLOGY

The proposed workflow is shown in Fig. 4. There are two major steps, i.e., 1) LCZ mapping fusing RSI and SVI data and 2) result analysis and interpretation. When interpreting the contributions of SVI for LCZ classification, two approaches are used in this study: Feature importance analysis and feature correlation analysis.

A. LCZ Mapping Fusing RSI and SVI

1) Generating Mapping Units: For LCZ mapping, the study area is first divided into 320 × 320 m² spatial grids. Then, using these uniform grids as basic mapping units, we extract the features for classifying the LCZ type of each mapping unit from corresponding Sentinel-2 imagery and SVIs. Specifically, each mapping unit is accompanied by a Sentinel-2 imagery patch cropped by its extent, as well as variable SVIs located within that mapping unit.

2) Extracting Features From RSI and SVI: For the RSI, each mapping unit includes a Sentinel-2 image patch with 32 × 32 pixel size. The statistics including mean, variance, maximum, minimum, skewness, and kurtosis of the pixel values of the ten bands are computed and stored as a 60-dimensional feature vector.

---

[1][Online]. Available: https://developers.google.com/maps/documentation/streetview/request-streetview
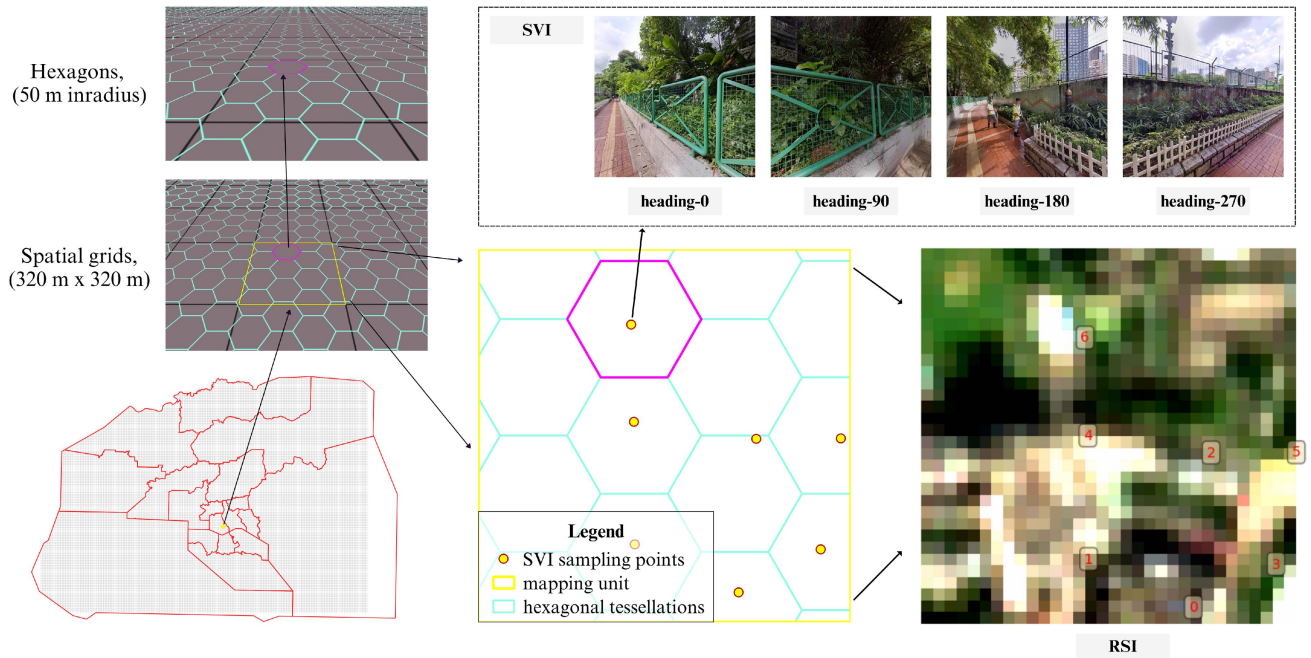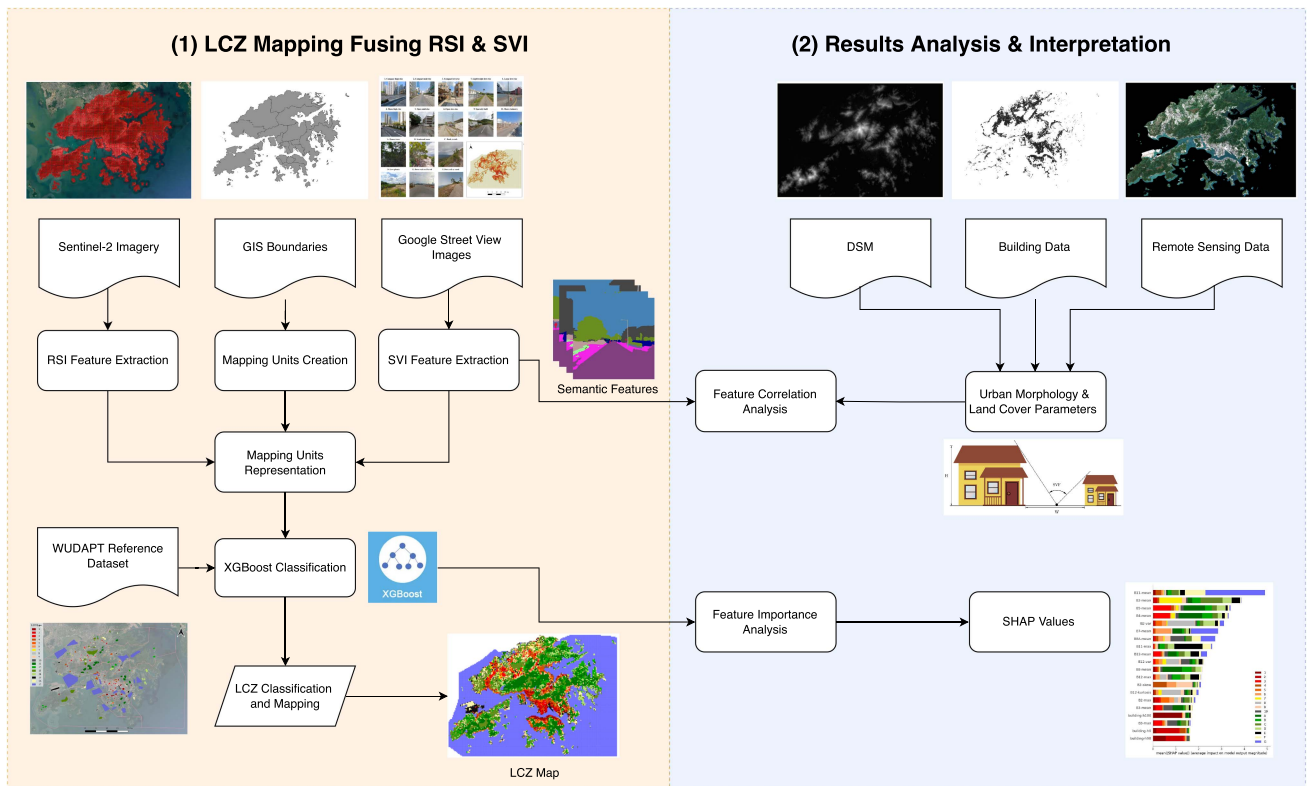
Fig. 3.   Illustration of study units.

Fig. 4.   Overview of the proposed workflow, including (1) LCZ mapping fusing RSI and SVIs and (2) result analysis and interpretation.

As for the SVIs, after the sampling process mentioned in Section III-B2, a spatial grid of 320×320 m^2 can contain variable SVI points. For each SVI point, the four SVIs are first fed into the pretrained DeepLab-v3+ neural network model [26] to classify the pixels as one of the 19 classes in the Cityscapes dataset [27], which is collected to understand the urban environment along roads. The 19 classes are road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, person, rider, car, truck, bus, train, motorcycle, and bicycle. Then, the pixel counts of the 19 classes are summed up as a

---

Algorithm 1: Fuse RSI and SVI Features for Mapping Units Representation.

Input: RSI feature set $\mathcal{F}_r = \{F_r^i\}_{i=1}^N$
Input: SVI feature set $\mathcal{F}_s = \{F_s^i\}_{i=1}^M$
Input: mapping unit set $\mathcal{U} = \{u_i\}_{i=1}^N$
Output: Representation set for mapping units
$\mathcal{F}_u = \{F_u^i\}_{i=1}^N$
1: **for** $u_i \in \mathcal{U}$ **do**
2:      $n_i = \text{GetNumOfSVI}(u_i)$
3:      **if** $n_i > 0$ **then**
4:          $F_u^i = \text{concat}(F_r^i, \text{aggregate}(\{F_s^j\}_{j=1}^{n_i}))$
5:      **else**
6:          $F_u^i = \text{concat}(F_r^i, \frac{\sum_{j=1}^M F_s^j}{M})$
7:      **end if**
8: **end for**

19-dimensional feature vector, and the four 19-dimensional feature vectors from four SVIs of the same SVI point are concatenated in the rowwise direction to finally obtain a 76-dimensional feature vector for each SVI point.

3) Classifying Mapping Units for LCZ Mapping: After the feature extraction process, the features of RSI and SVI are obtained. To fuse them together, Algorithm 1 is adopted. As for the RSI features, since each mapping unit only has a single RSI feature vector accompanied, no more operations are needed. For SVI features, limited by the spatial coverage of the street view points, some mapping units may contain no SVIs. To deal with this circumstance, the mean of all valid SVI features will be assigned to these mapping units. When there are multiple street view points located within a single mapping unit, the SVI features extracted from these SVI points are aggregated into one feature vector by using aggregation functions, such as mean and max pooling. After the process mentioned above, the RSI feature and the aggregated SVI feature can be concatenated to obtain the final fused feature vector to represent each mapping unit.

To perform the classification experiments, the rasterized WU-DAPT polygons are adopted as labels for the mapping units. The XGBoost model is chosen by reason of its high classification performance and interpretability. For the classification experiments, the cross-validation method with five folds is employed to attain more stable results. During each fold, the XGBoost model is fit on the partitioned training dataset, and the classification performance is evaluated on the test dataset. For the purpose of finding if fusing SVI actually helps improve the performance of classifying the LCZ, the classification experiment is performed on RSI, SVI, and fusing RSI and SVI together. For experiments using feature combinations of the three kinds mentioned above, the XGBoost model with the best performance is picked out as the classifier. Finally, for all the mapping units, the LCZ mapping results are obtained by predicting the LCZ type of each mapping unit with the classifier.

B. Feature Importance Analysis

In this study, SHapley Additive exPlanations (SHAP) analysis [28] is employed to examine the significance of the RSI and SVI features in predicting the LCZ types. SHAP analysis [28] is a method used to elucidate the predictions made by machine learning models. It assesses the influence of each feature on a particular prediction, enabling us to comprehend the factors that contribute positively or negatively to the output of the model. By aggregating the individual contributions of features across multiple predictions, it allows us to assess the overall importance of each feature in the performance of the model. Therefore, the SHAP analysis can help us identify RSI and SVI features that are consistently influential to LCZ predictions.

Specifically, since we adopted a fivefold cross-validation method to obtain more stable performance evaluation results, the contribution of each feature to the prediction is measured by calculating the mean absolute SHAP values on the full dataset. As part of this process, the XGBoost model was fitted on four folds of the full dataset for each fold. The mean absolute SHAP values were then calculated on these four folds of the full dataset. This calculation was repeated for all five folds of the cross validation. Therefore, the SHAP values were computed on each sample of the whole dataset four times, resulting in a total of 20 sets of SHAP values (5 folds $\times$ 4 sets of SHAP values). By aggregating the SHAP values, we can calculate the mean absolute SHAP values to measure the contributions of different features.

C. Feature Correlation Analysis

To conduct feature correlation analysis, we begin by computing the key LCZ indicators, which consist of SVF, BSF, ISF, and PSF. These indicators are calculated based on hexagons with a 50-m inradius. Each hexagon corresponds to an SVI point (as depicted in Fig. 5). The specific calculation methods for LCZ indicators are summarized in Table I. Once the indicators are computed, we proceed to measure the correlations between the SVI features and the obtained key LCZ indicators using Pearson correlation coefficients.

1) Calculating SVF: In analyzing the correlation between SVIs and key LCZ indicators, the SVF is considered as an important indicator with potential significant correlation [14]. Two approaches are used in this study to compute SVF values: one based on DSM data and the other based on SVIs. The DSM-based approach yields a raster SVF map, which is useful for large-scale calculations, while the SVI-based approach provides SVF values at specific location points, requiring additional processing to derive an SVF map. In the feature correlation analysis (see Section V-C), the DSM-based approach is used to generate an SVF map for the entire Hong Kong region. In the case studies (see Section V-D), the SVI-based approach is used to calculate a single SVF value from a single street view panorama.

For the DSM-based approach, the SVF calculation function provided by SAGA-GIS software [29] is employed. This function operates on raster data and requires specifying the number of directions and search radius for computing the SVF value for each cell. The algorithm consists of several steps. Initially, it determines the number of cells to be considered based on the defined search radius. Subsequently, it calculates the height difference between the current cell and its surrounding cells.
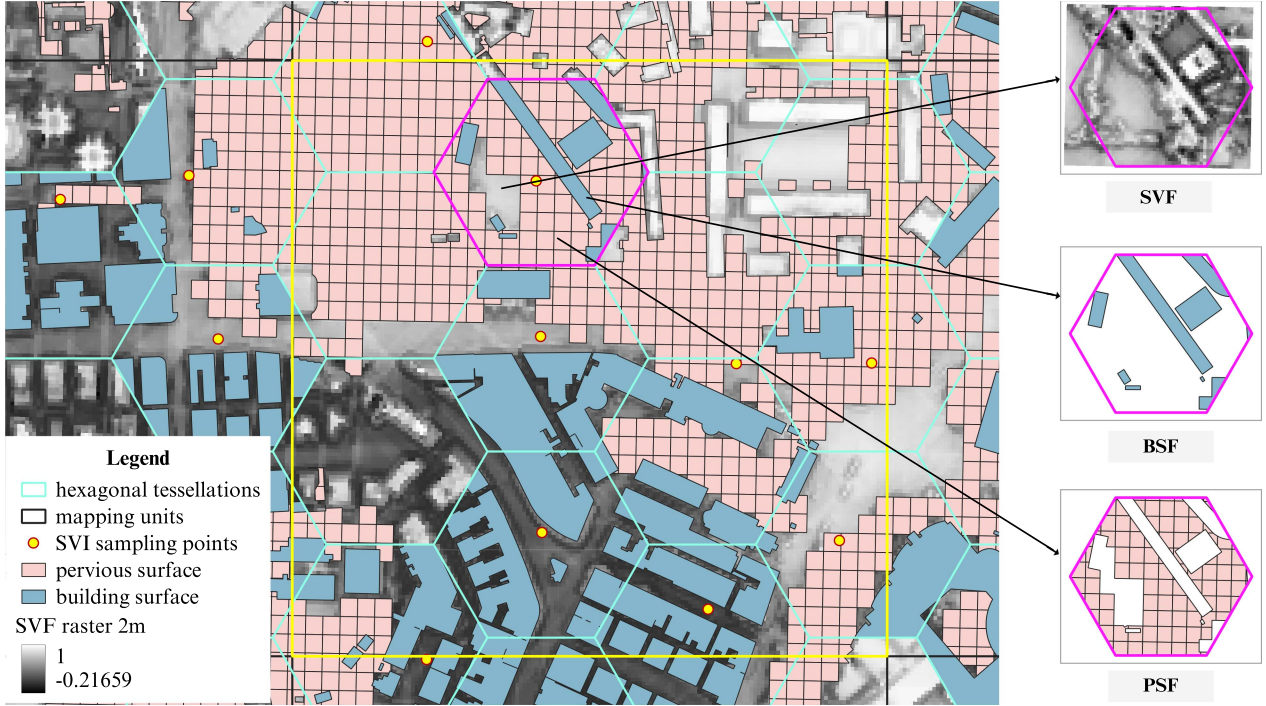
Fig. 5. Illustration of key LCZ indicators: SVF, BSF, and PSF.

TABLE I
DEFINITION OF KEY LCZ INDICATORS

Indicators	Definition	Data source	Calculation method
SVF	Areal mean SVF of nonbuilding areas $\sum_{i=1}^{n} S_i$	Building data and DSM data	$\text{SVF} = \frac{\sum_{i=1}^{n} S_i \cdot SVF_i}{\sum_{i=1}^{n} S_i}$
BSF	Building coverage ratio of the SVI sampling hexagon	Building data	$\text{BSF} = \frac{\sum_{i=1}^{n} BS_i}{S_{hex}}$
PSF	PSF of the SVI sampling hexagon	Remote sensing data	$\text{PSF} = \frac{\sum_{i=1}^{n} PS_i}{S_{hex}}$
ISF	ISF of the SVI sampling hexagon	Building and remote sensing data	$\text{ISF} = 1 - \text{BSF} - \text{PSF}$

Using this information, it determines the slope direction and slope angle of the current cell. Finally, it computes the SVF value using the derived slope direction and slope angle. Specifically, we input the 2-m DSM data into the aforementioned tool to derive a 2-m SVF map. Then, since each SVI is related to a hexagon area (as shown in Fig. 5), the zonal statistics method is applied to the SVF map to achieve a one-to-one relationship between SVI features and SVF values. The final SVF value of each hexagon is computed as the area-weighted average of the obtained 2-m SVF map

$$\text{SVF}_{\text{dsm}} = \frac{\sum_{i=1}^{n} S_i \cdot \text{SVF}_i}{\sum_{i=1}^{n} S_i} \qquad (1)$$

where n denotes the number of nonbuilding SVF grids (2-m resolution) that intersect with the hexagon, and S_i and SVF_i represent the area and the SVF value of the ith SVF grid, respectively.

The SVI-based approach aims to determine the proportion of sky pixels in the street-view panoramic images [30]. This is achieved by obtaining multiple SVIs of the same location, capturing a full 360° panoramic view, and stitching them together to generate panoramas. These panoramas are then processed using DeepLab-v3+ [26] pretrained on Cityscapes [27], which was described in Section IV-A. Through semantic segmentation, the pixels in the panorama are classified into one of the 19 Cityscapes classes. The upper half of the panorama is preserved, and the sky pixels are used to calculate the SVF value. To accomplish this, the upper half of the panorama is transformed from a cylindrical projection to an azimuthal projection, resulting in a fisheye image [31]. Then, the SVF can be computed based on the fisheye image using the following formula [32]:

$$\text{SVF}_{\text{svi}} = \frac{\pi}{2n} \sum_{i=0}^{n} \left(\sin \frac{\pi (2i-1)}{2n} \right) \left(\frac{p_i}{t_i} \right) \qquad (2)$$

where n represents the number of concentric annular rings of equal width which the fisheye image is divided into, while p_i and t_i are the number of the sky pixels and the total number of

pixels of the ring i. In our experiments (see Section V-D), n is set to 100 to obtain more precise results.

2) Calculating Surface Fraction Indicators: Surface fractions of different land cover types are also important LCZ indicators, such as BSF, PSF, and ISF.

Within each hexagon, the BSF value is computed by dividing the hexagon area by the building footprints area

$$\text{BSF} = \frac{\sum_{i=1}^{n} BS_i}{S_{\text{hex}}} \tag{3}$$

where n denotes the number of buildings that intersect with the hexagon area, BS_i denotes the footprint area of the ith building, and S_{hex} represents the area of the hexagon.

To compute PSF, bands of B4 (NIR) and B8 (RED) of the Sentinel-2 imagery are used to first obtain the NDVI value (NDVI $= \frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}}$); then, the Sentinel-2 imagery pixels with NDVI greater than 0.2 are filtered to compute the PSF values [6] within each hexagon

$$\text{PSF} = \frac{\sum_{i=1}^{n} PS_i}{S_{\text{hex}}} \tag{4}$$

where n denotes the number of filtered grids of the pervious map that intersect with the hexagon area, PS_i denotes the area of the ith grid, and S_{hex} represents the area of the hexagon.

Finally, the ISF is obtained by subtracting BSF and PSF

$$\text{ISF} = 1 - \text{BSF} - \text{PSF}. \tag{5}$$

3) Pearson Correlation Analysis: To interpret the correlation between the SVI features and the key LCZ indicators, the Pearson correlation coefficient is computed as a representation of the correlation strength. The Pearson correlation coefficient between two variables X and Y with n observations can be expressed using the following equation:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6}$$

where $r_{X,Y}$ is the Pearson correlation coefficient, x_i and y_i are the ith values of the two variables, and $\bar{x}$ and $\bar{y}$ are their respective means. The coefficient ranges from -1 to 1, where -1 and 1 represent a perfect negative and positive relationship, respectively, and 0 suggests no relationship. In addition, the p-value is calculated to determine if this correlation is statistically significant or simply due to chance. If the p-value is below a certain threshold (typically 0.05), it indicates a statistically significant correlation; conversely, if the p-value exceeds the threshold, the correlation is considered statistically insignificant.

Using the aforementioned calculation methods for the key LCZ indicators, for each SVI point, four LCZ indicators (SVF, BSF, ISF, and PSF) can be calculated. The Pearson correlation coefficient, along with the corresponding p-value, can be computed separately for each SVI feature and each LCZ indicator using all the available SVI samples.

TABLE II
OVERALL CLASSIFICATION RESULTS

Method	OA	OA$_{\text{urb}}$	OA$_{\text{nat}}$	WA	AA	Kappa
SVI	0.5340	0.4494	0.6343	0.8754	0.3598	0.4762
RSI	0.7926	0.4083	0.8669	0.9424	0.4510	0.7032
RSI+SVI	**0.8027**	**0.4517**	**0.8706**	**0.9470**	**0.4894**	**0.7175**

OA: overall accuracy, OA$_{\text{urb}}$: overall accuracy among the urban LCZ types, OA$_{\text{nat}}$: overall accuracy among the natural LCZ types, WA: weighted accuracy, and AA: average accuracy.

V. RESULTS AND ANALYSIS

A. Classification and Mapping Results

1) Overall Quantitative Results: To evaluate the classification results, the overall accuracy (OA), accuracy among the urban LCZ types (OA$_{\text{urb}}$), accuracy among the natural LCZ types (OA$_{\text{nat}}$), average accuracy (AA), weighted accuracy (WA), and Kappa coefficient are computed. The overall classification results of using different input data sources are presented in Table II.

From the table, we can see that the SVI classification has achieved a relatively good performance with OA over 50%, which shows that the street-level images alone already include some useful information associated with LCZ types. In contrast, RSI can achieve an accuracy of more than 70%. The fusion of RSI and SVI can further improve the classification performance in terms of all the evaluation metrics, particularly for classwise metric AA with nearly a 4% increase, which indicates that the addition of SVIs can help improve the classification of some difficult-to-recognize categories.

It is interesting to note that OA values lie in the middle of OA$_{\text{urb}}$ and OA$_{\text{nat}}$ (the accuracy of urban and natural LCZ types), and the classification results of natural types are significantly higher than that of urban types. This implies that urban built-up landscapes are more complex and difficult to recognize than the natural ones. However, SVI significantly improves OA$_{\text{urb}}$ over using RSI alone, which meets our expectation since street-level images include information on urban structures that are related to the LCZ scheme.

2) Class-Aware Quantitative Results: To further analyze the classification performance of different LCZ categories, the confusion matrices of different input data are shown in Fig. 6. From Fig. 6(a), we can find that the SVI is better at classifying natural LCZ types than built types, while from the perspective of categories, it has strong ability to classify class A (dense trees), class C (bush and scrub), and class G (water) and also performs well in classifying classes 1 (compact highrise), 2 (compact midrise), 3 (compact lowrise), 4 (open highrise), and class E (bare rock or paved); the remaining categories are generally challenging to correctly classify. From Fig. 6(b), we can see that the classification shows a similar result to the SVI, as RSI is also better at classifying natural types, which both indicate that the complex urban morphology contained in built types raises the difficulty of distinguishing them. On the other hand, the overall classification accuracy of RSI is superior to the SVI in the majority of categories, with the exception of class 4, where the SVI performs significantly better than RSI. Despite
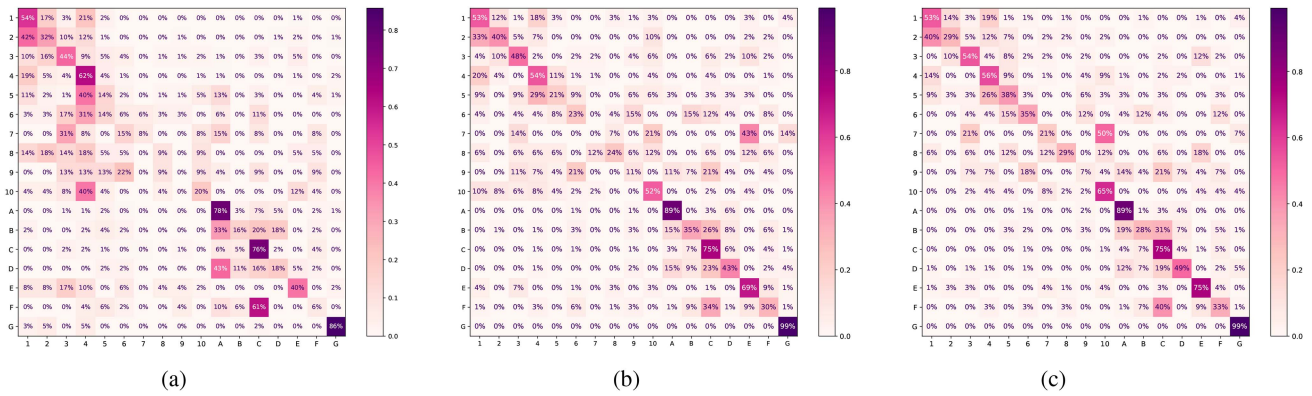
Fig. 6. Normalized confusion matrices of classification results using different input data. (a) Using SVI only. (b) Using RSI only. (c) Fusing SVI and RSI data.
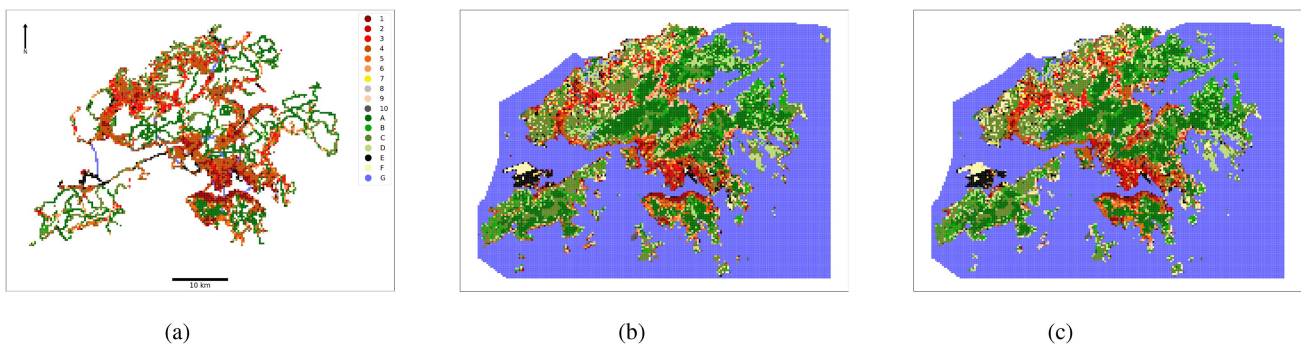

Fig. 7. LCZ mapping results using different input data. (a) Using SVI only. (b) Using RSI only. (c) Fusing SVI and RSI data.

this, the fusion of the SVI does help to improve the class-aware classification accuracy, as shown in Fig. 6(b) and (c), where the fused classification results in classes 3 (compact lowrise), 5 (open midrise), 6 (open lowrise), 7 (lightweight lowrise), and 10 (heavy industry) significantly outperformed the RSI. The fact that categories with low classification accuracy are dramatically enhanced after fusion more than categories with high classification accuracy is an intriguing finding that suggests the benefits of the fusion approaches.

3) Qualitative Mapping Results: For LCZ mapping, the classifier with the best performance of the five folds is picked out. The mapping results are shown in Fig. 7. As can be seen, RSI can generate a sensible LCZ map that is consistent with previous studies [15]. SVI data only cover limited areas along road networks, and the mapping results are generally reasonable. When fusing SVI with RSI, the mapping results can be further enhanced.

B. Feature Importance Analysis

SHAP analysis provides us with a reliable way to compare the importance of different features that contribute to the classification models. For the classification experiments using different input data, the mean absolute values of the SHAP values for features are presented in Figs. 8 and 9.

As can be seen from Fig. 8, the top 20 most important features are presented (y-axis of each subfigure) in the descending order.

For each feature, its contributions to different target LCZ categories are denoted by different colors, which correspond to the colors used in LCZ mapping results (as shown in Fig. 7). For the SVI [see Fig. 8(a)], we can see that the features of the sky play the most important role in the classification task, accompanied by vegetation, road, and building, composing the top five most important features. For RSI [see Fig. 8(b)], the bands of B11 (SWIR 1), B2 (Blue), B4 (Red), and B8 (NIR) are the top five most important features for classifying the samples to certain LCZ types.

For fusing RSI and SVI [see Fig. 8(c)], we can find that among the top 20 most important features, only the feature of building from the SVI is reserved, the others are all supported by RSI. When focusing on the features of building-h180, building-h0, and building-h90 in Fig. 8(c), we can find that LCZ type 1 (compact highrise) and type 2 (compact midrise) account for the largest percentage, which means that in the classification task of fusing RSI and SVI, the features of building contribute mostly to classifying the urban types of compact highrise and compact midrise. Compared with the other features, we can also find that the contributions to LCZ type 1 and LCZ type 2 are almost all from the features of the building.

Fig. 9 offers a different perspective to analyze the feature importance. Specifically, for each LCZ type (indicated in the y-axis), the top 20 most important features that contribute to the classification are presented, which are denoted by different colors. From Fig. 9(a), for compact built types (classes 1–3), we
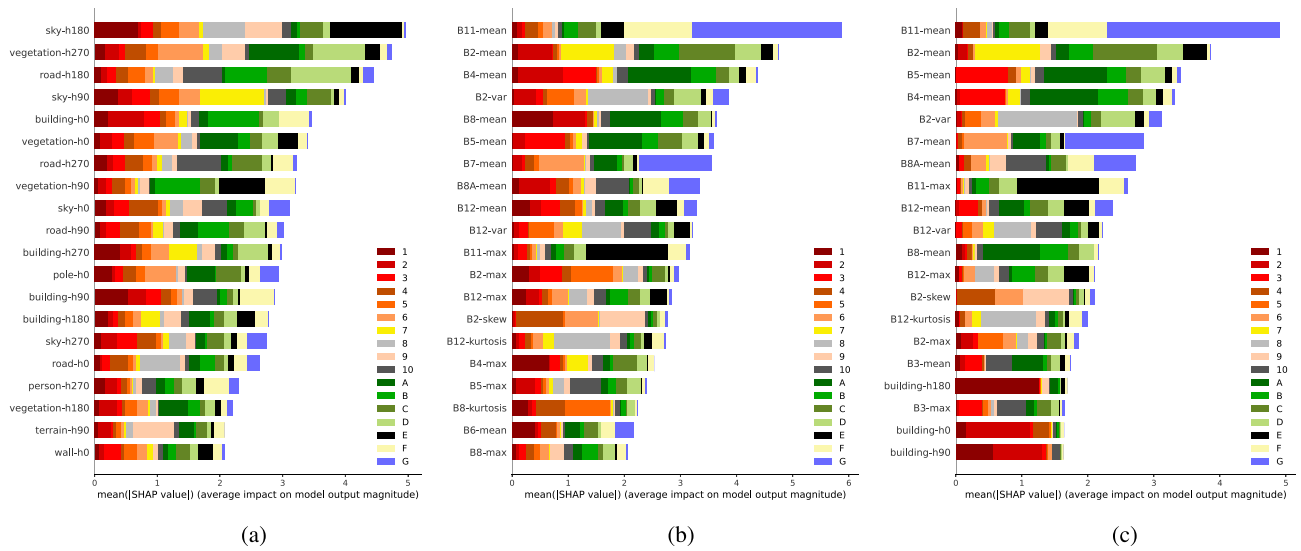
Fig. 8. SHAP summary plot for XGBoost models trained on different input data, with the top 20 most important features listed. (a) SVI. (b) RSI. (c) RSI+SVI.

can find that the features of the building play an important role in the classification task, while for the open built types (classes 4–6), we can find that nature-related features such as vegetation and sky contribute the most. From Fig. 9(b), for the compact built types, bands of B4 and B8 play a significant role in both class 1 (compact highrise) and class 2 (compact midrise), while class 3 (compact lowrise) actually benefits most from B5, which has a direct relationship with vegetation. This is consistent with the idea that as the height of the built types decreases, the effects of vegetation may gradually work. When focusing on the feature importance composition of LCZ type 1 and type 2 in Fig. 9(b) and (c), we can find that in Fig. 9(b), the features of B8 (NIR) and B4 (Red) occupy the largest part, while this important role of B8 and B4 is replaced by the features of building in Fig. 9(c), which implies that the building part of SVI features makes a vital contributions to the LCZ classification.

C. Feature Correlation Analysis

For correlation analysis, the Pearson coefficients between the SVI semantic features and four key LCZ indicators (SVF, BSF, PSF, and ISF) are calculated and presented in Fig. 10. For SVF, we can see that features of sky, terrain, and vegetation are the most positively correlated, while building features are the most negatively correlated. This is consistent with the definition of SVF, which is described as, "The fraction of the overlying hemisphere occupied by sky" [33]. Since buildings block the sky most from the view on the ground, the correlation between SVF and the building feature appears strongly negative.

For BSF, vegetation, sky, and terrain are the only semantic categories negatively correlated, while building dominates the positive correlation. Since BSF is computed as the ratio of the building footprint area and the total area of the hexagon, it is unsurprising to find that the feature of the building has the most strong positive correlation with the BSF over the other positively correlated features. On the contrary, for PSF, vegetation and terrain types are positively correlated. Specifically, vegetation

shows a very strong positive correlation, and terrain shows a relatively weak positive correlation, while all others are negatively correlated, with building, road, and sidewalk being the most negatively correlated semantic categories.

For ISF, only vegetation and terrain are negatively correlated, and the former shows a strong negative correlation, while the correlation of the latter is relatively weak. The features of sky, building, and road have positive competitive correlations with the ISF. While it is easy to understand that building and road lead the positively correlated categories since they contribute directly to most impervious surfaces, it is interesting to notice that sky also presents a strong positive correlation with ISF. One possible interpretation is that a more visible sky means fewer trees (included in the vegetation category), and fewer trees mean a more impervious surface.

The semantic features extracted from SVIs show strong correlations with the urban morphology/land cover parameters, which serve as defining criteria of the LCZ classification system. This suggests that the SVIs have the potential to contribute and enhance LCZ mapping, as they already contain essential information relevant to the LCZ definition.

D. Case Studies

In order to further investigate how SVIs contribute to improving LCZ classification, specific case studies are selected, and the classification results using RSI and fusing RSI and SVI are compared. In addition, the content of SVIs is also analyzed with details to help better understand what within SVI helps. Specifically, we select typical samples based on the observations from the confusion matrices shown in Fig. 6. We find that the prediction performance of class 5 (open midrise) and class 10 (heavy industry) has been improved the most by fusing RSI and SVI. Class 7 (lightweight lowrise) is not considered because the sample of class 7 is relatively rare in Hong Kong. Based on this finding, two case studies are conducted, and the location, corresponding RSI, and SVIs are presented in Fig. 11. Specifically,
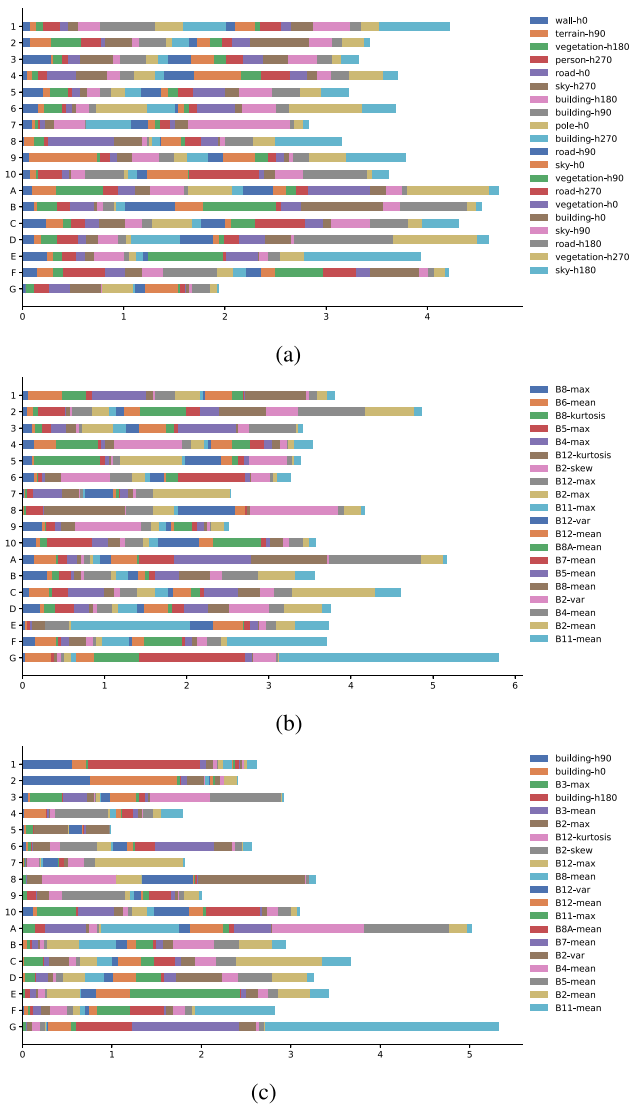
Fig. 9. Transposed SHAP summary plot for analyzing the feature importance of each LCZ type. (a) SVI. (b) RSI. (c) RSI+SVI.

both the street-view panoramas and the corresponding computed SVF values are leveraged to probe into details.

The first study unit A lies in the Southern District of Hong Kong. The true type of this sample is class 5 (open midrise), while when predicted with the XGBoost model trained on the RSI dataset, the predicted type is class 3 (compact lowrise). As shown in Fig. 11, there are nine SVI points lying in the study area in case A. From these SVIs, we can see that vegetation and sky occupy the largest areas, and all the scenes in the SVIs look open and not compact. The second study unit B lies in the Kwai Tsing District, with the ground truth of class 10 (heavy industry). When predicted with the model trained on the RSI dataset, the incorrect predicted type is class 2 (compact midrise). In the study area, seven SVIs are presented within the spatial extent. From the third and sixth SVI, we can find that some cylindrical tanks, which are mostly used in heavy industry factories, are shown in the left part. These clues may be the key to help classify sample B as class 10 (heavy industry).

The SVF values of the two sample regions can be computed based on (2). For sample A, the final mean SVF value of the nine panoramas is 0.518, while for sample B, the mean SVF value is 0.737. From the definitions of the LCZ types, the SVF value ranges of different LCZ types can be found, which are 0.3–0.6 (class 2, compact midrise), 0.2–0.6 (class 3, compact lowrise), 0.5–0.8 (class 5, open midrise), and 0.6–0.9 (class 10, heavy industry) [2]. We can see that there is some overlap between different classes, which suggests that SVF values alone are not always sufficient enough to distinguish all the LCZ classes. In our case A, the SVF value of 0.518 obtained from the panoramas cannot sufficiently distinguish class 5 from class 3. One potential interpretation is that the vegetation of class 5 has a similar effect of blocking the sky as the buildings of class 2. Thus, the range of SVF values of class 3 and class 5 has a certain amount of overlap. For sample B, the SVF value of 0.737 is high enough and clearly suggests that it should be classified as class 10.

VI. DISCUSSION

A. Availability and Quality of Street-Level Images

1) Data Availability: To use street-level images for practical LCZ mapping, data availability is critical. There are several different sources of street-level images. The most commonly used street-level images are acquired from online street view services. In addition to Google Street View, there are also other similar SVI-based services provided by Baidu,[2] Tencent,[3] etc., which have relatively limited coverage than Google and only cover cities in China. Besides map service providers, there is a crowdsourcing street-level image platform named Mapillary[4] on which users can upload and share geo-tagged street-level images. Street-level images can also be accessed through social media platforms, such as Twitter, Flickr, etc. Furthermore, street-level images can also be captured by users themselves according to customized needs. Diverse sources of SVIs facilitate their practical use in LCZ mapping.

2) Data Quality: Despite the various sources of street-level images, there are some inherent issues with data quality. The most prominent limitation of SVIs is the limited spatial and temporal coverage and resolution. SVI points are distributed along roads, which are normally sparsely distributed. Besides, most SVIs are captured in urban areas. In addition to spatial coverage, the update of SVIs is difficult to control, which is determined by the street view service providers or sharing users. The street scenes may change as time goes by, and the update frequency is highly uncertain across different regions. All these limitations on spatial and temporal coverage have significantly limited the quality of SVIs for LCZ mapping. Studies on how to leverage limited SVIs to enhance the LCZ mapping results are worthy of further efforts.

Besides, the position accuracy of SVIs is also questionable. Unlike remote sensing images that usually have very accurate location information, the positions of street-level images are

[2][Online]. Available: https://map.baidu.com/
[3][Online]. Available: https://map.qq.com/
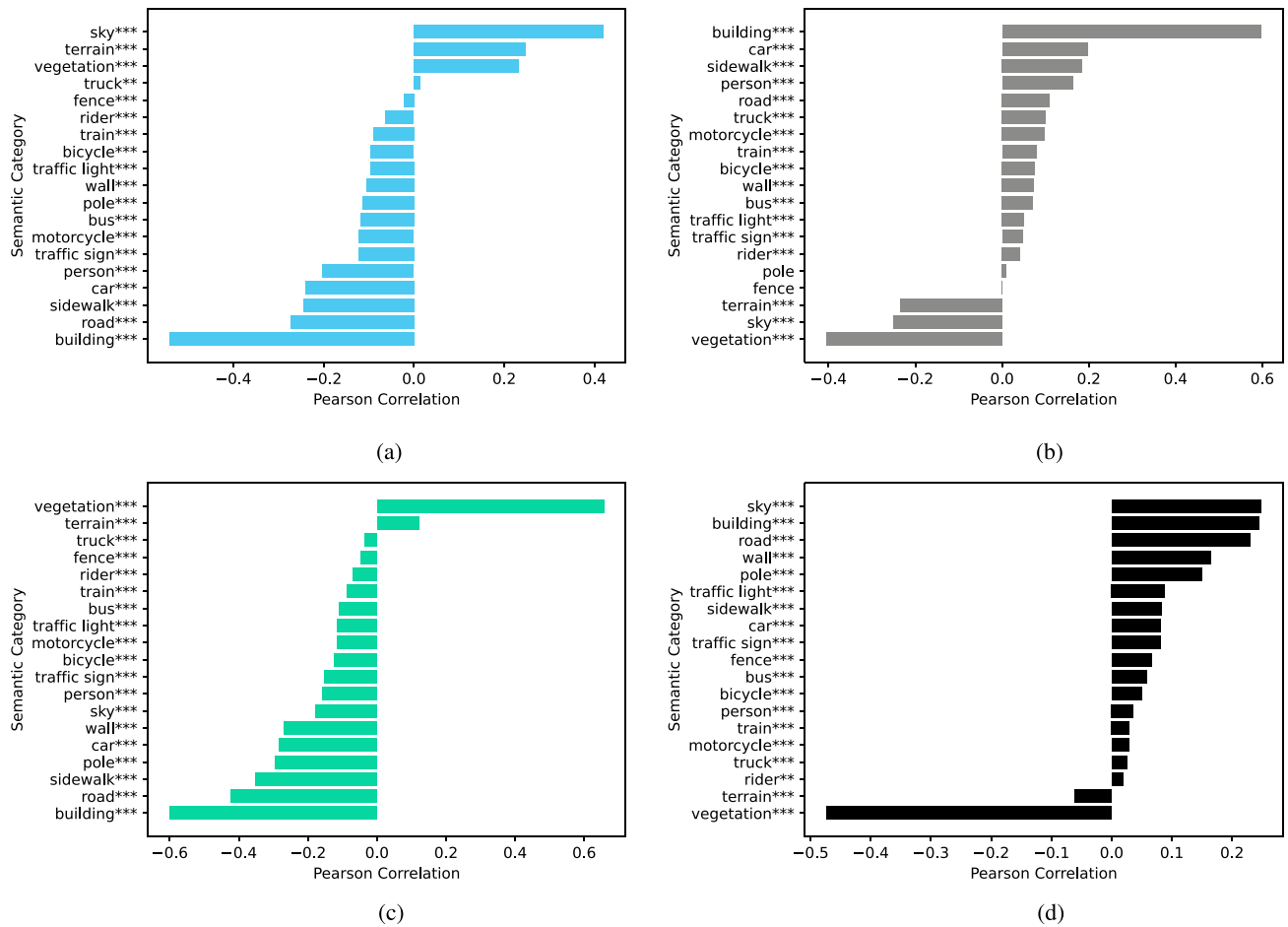[4][Online]. Available: https://www.mapillary.com/

Fig. 10.   Feature correlation analysis between SVI semantic categories and key LCZ indicators. (a) SVF. (b) BSF. (c) PSF. (d) ISF (***: $p < 0.001$, **: $p < 0.01$, and *: $p < 0.05$).

usually coarse and sometimes even incorrect due to the positioning device and/or the online distribution mechanism. This issue is also not unusually seen for crowdsourcing platforms and social media platforms, where the geotags are usually questionable. Therefore, it is important to know the existing location quality issues of the SVIs in order to circumvent inappropriate operations and analysis.

B. Limitations and Future Directions

Although the study explored how street-level images can help with LCZ mapping and revealed some interesting findings, there are several limitations that deserve further research in the future.

First, to make it easier to interpret the contributing factors, only spectral features and statistics of semantic segmentation features are extracted from RSI and SVI, respectively. The extracted features are relatively straightforward and simplified and are far away from exhaustive, which facilitates the understanding and interpretation but may also limit the accuracy of the classification models. How to better balance the tradeoff is worth further study.

Second, it is worth noting that not all the semantic segmentation classes in the Cityscapes dataset are directly related to

the components of the LCZ scheme. This may explain why only a few of the extracted semantic features from the SVI have shown significant contributions in improving LCZ mapping performance. Furthermore, this also affects the feature correlation analysis since the semantic features are employed as proxy and are not directly related to LCZ indicators (except for SVF, which can be calculated directly from the semantic segmentation results). To address this issue, future studies should explicitly consider LCZ-related information when determining the categories into which pixels should be classified. This would help to improve both LCZ mapping performance and the accuracy of feature correlation analysis.

In addition, in this study, Hong Kong is selected as the study area. However, it will be interesting to include more urban areas of different types to examine the influence of street-level images on them, which may give us more findings.

VII. CONCLUSION

Timely and accurate LCZ maps play a crucial role in urban climate studies, particularly in understanding phenomenon like UHIs. Remote-sensing-based LCZ mapping is becoming popular, and street-level images can serve as a complementary data
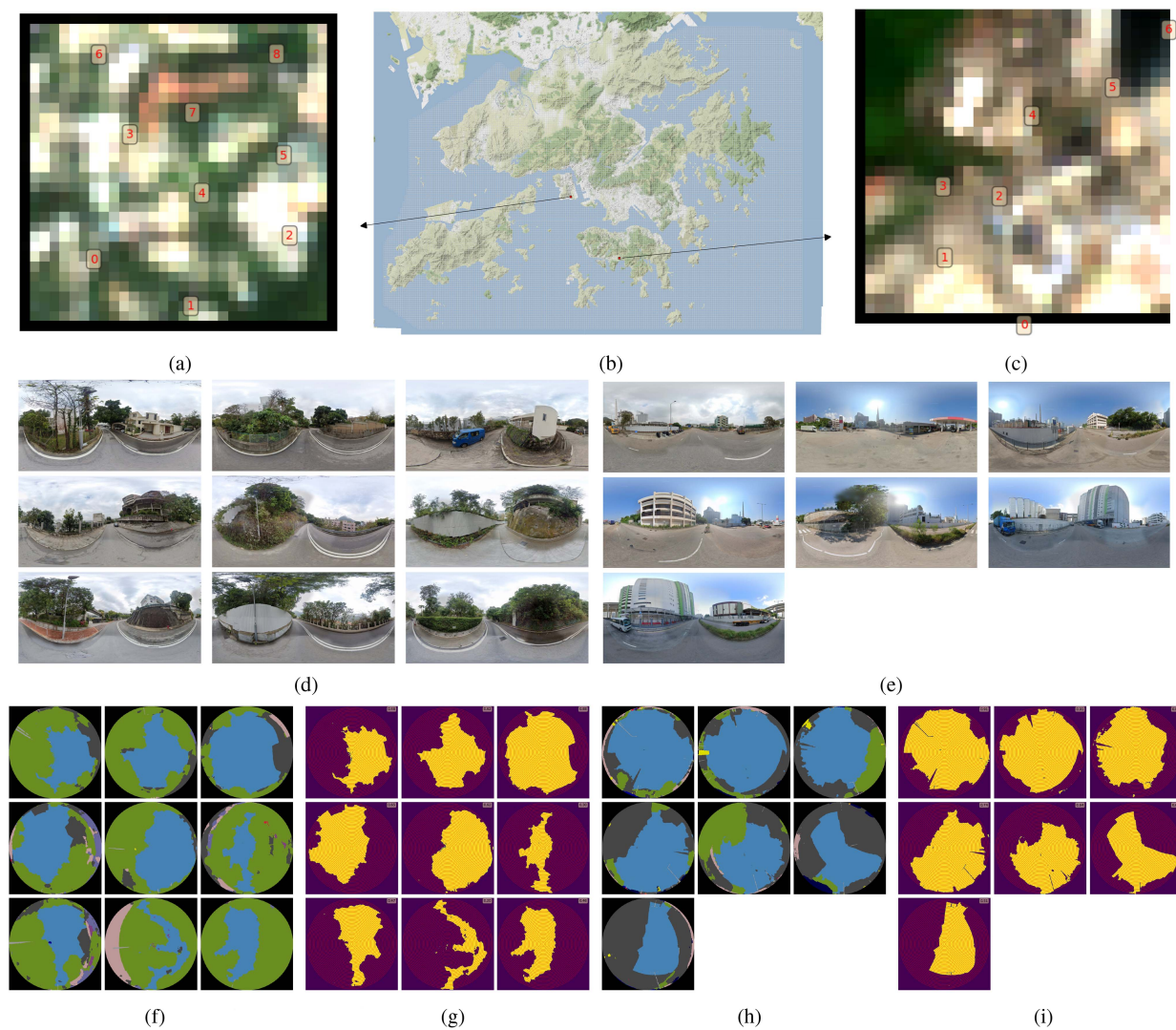
Fig. 11. Overview of the case studies. (a) and (c) RSI image patches with annotations indicating the locations of SVI points. (b) Study area overview. (d) and (e) Corresponding SVIs. (f) and (h) Semantic segmentation results of fisheye images transformed from the SVIs (blue: Sky, green: Vegetation and trees, and gray: Buildings). (g) and (i) SVF calculation results using annular rings.

source to further enhance the results by providing ground-level details. In this study, our primary objective was to explore the potential of street-level images in improving remote-sensing-based LCZ mapping. For this purpose, we selected Hong Kong as our study area and conducted extensive experiments that combined remote sensing images with street-level images for LCZ classification and mapping. The findings of our study demonstrated that the incorporation of street-level images can enhance the accuracy of the classification results. To delve deeper into the underlying mechanisms behind this improvement, we conducted feature importance analysis and feature correlation analysis. The analysis of feature importance revealed the substantial contribution of building features to the improvement of LCZ classification results, especially for compact urban types such as compact highrise and compact midrise. These findings indicate that the building information captured in SVIs has a positive effect on the mapping process. In addition, the correlation analysis highlighted strong associations between SVI features and

key LCZ indicators, further reinforcing the relevance of SVIs in LCZ classification.

In the future, several directions can be pursued to advance the field. First, it is essential to investigate how to strike a balance between the tradeoff of SVI feature complexity and the interpretation of SVI contributions, ensuring the optimal utilization of these images. Second, exploring ways to incorporate LCZ-related information when determining the semantic categories of SVIs can lead to improved LCZ mapping performance. Finally, expanding the study to include additional study areas would offer a more comprehensive understanding of the impact of street-level images on LCZ mapping.

REFERENCES

[1] A. M. Rizwan, L. Y. Dennis, and L. Chunho, "A review on the generation, determination and mitigation of urban heat island," *J. Environ. Sci.*, vol. 20, no. 1, pp. 120–128, 2008.

[2] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 12, pp. 1879–1900, 2012.

[3] J. Xue, R. You, W. Liu, C. Chen, and D. Lai, "Applications of local climate zone classification scheme to improve urban sustainability: A bibliometric review," *Sustainability*, vol. 12, no. 19, 2020, Art. no. 8083.

[4] R. Cao et al., "Integrating satellite and street-level images for local climate zone mapping," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 119, 2023, Art. no. 103323.

[5] G. Thomas, A. P. Sherin, S. Ansar, and E. J. Zachariah, "Analysis of urban heat Island in Kochi, India, using a modified local climate zone classification," *Procedia Environ. Sci.*, vol. 21, pp. 3–13, 2014.

[6] Y. Zheng et al., "GIS-based mapping of local climate zone in the high-density city of Hong Kong," *Urban Climate*, vol. 24, pp. 419–448, 2018.

[7] R. Wang, C. Ren, Y. Xu, K. K.-L. Lau, and Y. Shi, "Mapping the local climate zones of urban areas by GIS-based and WUDAPT methods: A case study of Hong Kong," *Urban Climate*, vol. 24, pp. 567–576, 2018.

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

[9] X. X. Zhu et al., "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 3, pp. 76–89, Sep. 2020.

[10] C. Ren et al., "Assessment of local climate zone classification maps of cities in China and feasible refinements," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 18848.

[11] F. Biljecki and K. Ito, "Street view imagery in urban analytics and GIS: A review," *Landscape Urban Plan.*, vol. 215, 2021, Art. no. 104217.

[12] Y. Yan and B. Huang, "Estimation of building height using a single street view image via deep neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, pp. 83–98, 2022.

[13] G. Xu, X. Zhu, N. Tapper, and B. Bechtel, "Urban climate zone classification using convolutional neural network and ground-level images," *Prog. Phys. Geography: Earth Environ.*, vol. 43, no. 3, pp. 410–424, 2019.

[14] M. Ignatius et al., "Local climate zones: Lessons from Singapore and potential improvement with street view imagery," *ISPRS Ann. Photogrammetry, Remote Sens., Spatial Inf. Sci.*, vol. X- 4/W2-2022, pp. 121–128, 2022.

[15] Y. Zheng et al., "GIS-based mapping of local climate zone in the high-density city of Hong Kong," *Urban Climate*, vol. 24, pp. 419–448, 2017.

[16] S. Liu and Q. Shi, "Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China," *ISPRS J. Photogrammetry Remote Sens.*, vol. 164, pp. 229–242, 2020.

[17] R. Cao et al., "Integrating aerial and street view images for urban land use classification," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1553.

[18] R. Cao and G. Qiu, "Urban land use classification based on aerial and ground images," in *Proc. IEEE Int. Conf. Content-Based Multimedia Indexing*, 2018, pp. 1–6.

[19] J. Kang, M. Körner, Y. Wang, H. Taubenböck, and X. X. Zhu, "Building instance classification using street view images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 44–59, 2018.

[20] S. Law, B. Paige, and C. Russell, "Take a look around: Using street view and satellite images to estimate house prices," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, 2019, Art. no. 54.

[21] Y. Yan and Y. Ryu, "Exploring Google street view with deep learning for crop type mapping," *ISPRS J. Photogrammetry Remote Sens.*, vol. 171, pp. 278–296, 2021.

[22] F. Zhang, L. Wu, D. Zhu, and Y. Liu, "Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns," *ISPRS J. Photogrammetry Remote Sens.*, vol. 153, pp. 48–58, 2019.

[23] Y. Kang, F. Zhang, S. Gao, H. Lin, and Y. Liu, "A review of urban physical environment sensing using street view imagery in public health studies," *Ann. GIS*, vol. 26, pp. 261–275, 2020.

[24] G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Comput., Environ. Urban Syst.*, vol. 65, pp. 126–139, 2017.

[25] F. Huang et al., "Mapping local climate zones for cities: A large review," *Remote Sens. Environ.*, vol. 292, 2023, Art. no. 113573.

[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 833–851.

[27] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.

[28] M. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[29] J. Böhner, K. R. McCloy, and J. Strobl, *SAGA—Analysis and Modelling Applications*. Göttingen, Germany: Goltze, 2006.

[30] Y. Xia, N. Yabuki, and T. Fukuda, "Sky view factor estimation from street view images based on semantic segmentation," *Urban Climate*, vol. 40, 2021, Art. no. 100999.

[31] X. Li, C. Ratti, and I. Seiferling, "Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google street view," *Landscape Urban Plan.*, vol. 169, pp. 81–91, 2018.

[32] L. Chapman, J. E. Thornes, and A. V. Bradley, "Rapid determination of canyon geometry parameters for use in surface radiation budgets," *Theor. Appl. Climatol.*, vol. 69, no. 1, pp. 81–89, 2001.

[33] T. R. Oke, "Canyon geometry and the nocturnal urban heat island: Comparison of scale model and field observations," *Int. J. Climatol.*, vol. 1, no. 3, pp. 237–254, 1981.

Cai Liao is working toward the master's degree with the School of Geodesy and Geomatics, Wuhan University, Wuhan, China, majoring in Cartography and Geographic Information Engineering.

He is also a Research Assistant with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong. His research interests include GeoAI and urban informatics.
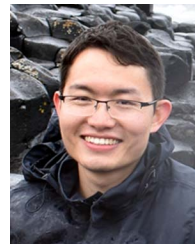
Rui Cao (Member, IEEE) received the B.Eng. and M.Eng. degrees from Wuhan University, Wuhan, China, and the Ph.D. degree from the University of Nottingham, Nottingham, U.K.

He is currently a Research Assistant Professor with the Department of Land Surveying and Geo-Informatics and the Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong. His research interests include geographic information science and particularly interested in GeoAI and urban informatics, with an ultimate goal of contributing to sustainable and human-centered smart cities for the benefit of society.

Qi-Li Gao received the Ph.D. degree in geographic information science from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

She is currently a Research Fellow with the Centre for Advanced Spatial Analysis, University College London, London, U.K. Her research interests include spatial–temporal data mining, urban informatics, and analysis on human mobility and social inequality toward sustainable urban and transport planning.

Jinzhou Cao received the Ph.D. degree in geographic information science from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

He is currently an Assistant Professor with the College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China. His research interests include urban big data mining, GeoAI, and urban analytics on the interaction between human mobility and urban environments.

Nianxue Luo received the Ph.D. degree from Wuhan University, Wuhan, China, in 2002.

He is currently a Professor with the School of Geodesy and Geomatics, Wuhan University. His research interests include the geographic information science project, disaster monitoring, emergency management, and software design for surveying and mapping data process.