Coupling Dual Graph Convolution Network and Residual Network for Local Climate Zone Mapping

Yedan Yu, Jie Li ⓘ, *Member, IEEE*, Qiangqiang Yuan ⓘ, *Member, IEEE*, Qian Shi ⓘ, *Senior Member, IEEE*, Huanfeng Shen ⓘ, *Senior Member, IEEE*, and Liangpei Zhang ⓘ, *Fellow, IEEE*

Abstract—**Local climate zone (LCZ) has become a new standard classification scheme in urban landscapes and showed great potential in urban climate research. Traditional classifiers and ordinary neural networks only consider the spectral or local spatial features of the pixel, ignoring the effect of nonlocal information on the LCZ classification. The graph convolutional network (GCN) has been used to exploit the relationship between adjacent and global land covers owing to the ability to conduct flexible convolution over graphs. In this work, we integrated a convolutional neural network and two GCNs into an end-to-end hybrid framework and generated LCZs directly from the original images. Local-, regional-, and global-level features were extracted and grouped complementarily to foster better performance. Experiments were conducted in six cities around the world to verify the effectiveness of our method. Results showed that the average classification accuracy of the six cities reached 0.956 and performed better than any other comparable model. Ablation experiments also demonstrated the mutual promotion of the different modules. Finally, the small sample experiment provided a practical reference for the LCZ classification in the absence of samples in future.**

Index Terms—**Graph neural network, local climate zone (LCZ), urban climate.**

I. INTRODUCTION

CITIES are currently home to half of the world's population, and all of the projected 1.1 billion global population growth will be virtually in urban areas over the next 15 years [1]. More severe challenges are expected to occur in urban ecological areas, but progress in urban climate science is severely restricted by the lack of useful information on the form and function of cities at a high spatial resolution [2]. A climate-related global urban classification scheme is essential. Many classification

Yedan Yu, Jie Li, and Qiangqiang Yuan are with the School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China (e-mail: yyd2345@gmail.com; aaronleecool@whu.edu.cn; qqyuan@sgg.whu.edu.cn).

Qian Shi is with the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510085, China (e-mail: shixi5@mail.sysu.edu.cn).

Huanfeng Shen is with the School of Resource and Environmental Science, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

Liangpei Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3132394

models are available, such as the national land cover dataset product of the United States, which has 20 land cover types with only four urban classes based on urban building density [3]; the European CORINE (i.e., Co-ORdinated INformation on the Environment) land cover database, which contains 44 land cover categories with only three urban areas [4]; and the Global Land Cover product (GLC30) [5] with 10 classes and only "artificial surface" referring to all artificial impervious surfaces. These classification schemes only have rough city types and do not provide the necessary detail. Moreover, some are designed for specific cities and cannot be generalized to global classifications. In addition, these models do not take into account climate-related attributes. For this reason, the local climate zone scheme was proposed to fulfill all the requirements [6].

The LCZ is a universal, culturally neutral, climate-based classification scheme. It is defined as regions of uniform surface cover, structure, material, and human activity that span hundreds of meters to several kilometers horizontally. As shown in Fig. 1, there are 17 standard types in the LCZ scheme, including "built types" LCZ 1–10 on the left and "land cover types" LCZ A–G on the right. Each climate zone has a unique set of physical attributes, including sky view factor, aspect ratio, building surface fraction, etc. These attributes are climate-related and influence the overlying near-surface atmosphere. Each property is measurable and nonspecific as to place or time, making the LCZ scheme globally transferable and comparable. Last, LCZ classifies cities in a more detailed way compared with other classification systems. Therefore, LCZ maps can be useful in studies about the configuration of cities and their impact on human and environmental applications, such as the heat island effect [7] and the effects of respiratory particulate matter [8].

To obtain satisfactory classification performance, many methods have been proposed. The LCZ classification using remote sensing images has become one of the mainstream methods. With freely available satellite imagery, an automated LCZ mapping procedure was developed [2]. After that, traditional classifier and convolutional neural network (CNN) were adopted widely in this field. However, all of these methods focus on local information. The role of regional and global features in LCZ has not been explored. Due to the strong scale dependence on the LCZ classification, the same scene may be classified into different types at diverse scales. For example, Phoenix Park was classified as LCZ B in 1000-m resolution, but it was classified as LCZ5, A, B, C, D, and F in 200-m resolution [2]. A larger scale can help in maintaining the consistency of the classification and
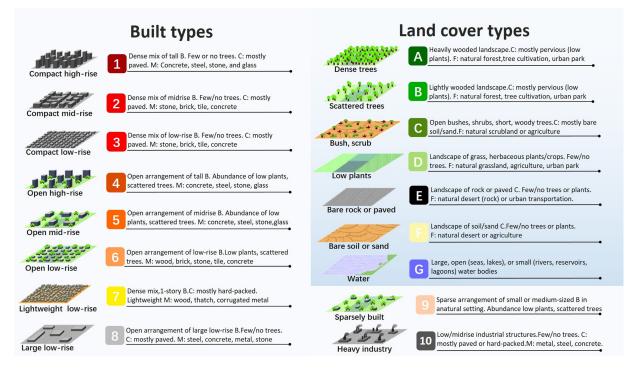
Fig. 1.   LCZ scheme and the characteristics of 17 types. The subfigures were modified from [6]. B: Buildings; C: Cover; M: Materials; F: Function.

avoid a fragmented classification map. Hence, a region-based graph convolutional network is proposed in this work. It helps to obtain a larger receptive field and consider neighborhood textures comprehensively by connecting adjacent superpixels. Furthermore, a global GCN is designed to connect the same LCZ type over a long distance in the global scope. There are edges between one node and any other nodes and the weight of these edges update based on the similarity of nodes. So each node can be updated with global information. Therefore, a three-branch network is constructed in consort with the local information generated by CNN. The main contributions of this article are as follows.

1) A novel hybrid framework is designed by integrating a CNN with two GCNs. To our knowledge, this study is the first to introduce GCN to classify LCZ maps. In contrast with traditional LCZ classification methods, the entire original image can be classified by combining local-, regional-, and global-level features from three branches without additional cutting and feature extraction.

2) A dual GCN structure is proposed to break through the limits of local information. The regional GCN (RGCN) branch can deeply excavate correlations between larger-scale land covers and generate a region-consistent classification map without many fragments. Meanwhile, the global GCN (GGCN) links units over the global area to capture the intra- and interclass relationship between LCZ types.

3) The proposed network demonstrates its validity and superiority in the experiments by incorporating the advantages of CNN and GCN. Furthermore, when the sample size of a city is insufficient, the proposed method may achieve satisfying results by using a small sample classification

strategy rather than employing the unstable sample transfer strategy.

The rest of this article is organized as follows: The related work is introduced in Section II. The study area and data are described in Section III. The proposed method and GCN-related knowledge are introduced in Section IV. The qualitative and quantitative results of the experiments are illustrated in Section V. The discussions are exhibited in Section VI. Finally, the work is summarized and concluded in Section VII.

II. RELATED WORK

A. Traditional Classifier

The World Urban Database Access Portal Tools (WUDAPT) project [2] has adopted the LCZ scheme as a basic description of urban land-cover. It has proposed a universal LCZ mapping process for automatically classifying urban areas into LCZ types. Landsat-8 and software (SAGA and Google Earth) were applied to establish a simple manufacturing workflow that someone without specialist knowledge can participate in. This approach greatly expanded the influence and application of LCZ. However, the average overall accuracy (OA) of the 90 LCZs uploaded on the WUDAPT portal was 74.5%, leaving much room for improvement [9].

A variety of data was added to this framework. Synthetic aperture radar (SAR) data and gray-level co-occurrence textures were included in WUDAPT for better performance [10]. Open Street Map (OSM) was also used to enhance the development of LCZ maps, and improvements can be achieved for certain classes [11].

In the WUDAPT operation process, it is the random forest classifier embedded in SAGA software that works for

classification. Using these classifiers alone can get the same effect, and the operation is more flexible and convenient, which is conducive to the free combination of features. Thus, four classifiers [Naïve Bayes, multilayer perceptron, support vector machine, and random forest (RF)] were adopted to test nine different combinations of features of Landsat or SAR or both datasets. The result showed that the combination of these datasets provides some improvement [10]. In the 2017 IEEE Geoscience and Remote Sensing Society Data Fusion Contest, the algorithm based on decision tree ensemble classifiers won first place [12]. Many features were derived from Landsat 8 and OSM data, including spectral reflectance, spectral indices, OSM features, and spatial features. It reached the highest accuracy of 74.94%. However, these features and indices from the original image and auxiliary data needed tedious feature extraction preprocessing.

B. Convolutional Neural Network

With the advances of deep learning models especially CNNs, powerful feature representations can be extracted automatically [13]. CNN has shown exemplary performance in various computer vision tasks [14]–[18]. In remote-sensing image scene classification, CNNs learn discriminative features with small within-class scatter but significant between-class separation, which significantly improved the classification performance [19]. Thus, many research works have developed CNN-based methods for the LCZ classification. Rosentreter *et al.* [20] derived high-resolution LCZs using multitemporal Sentinel-2 composites and a six-layer CNN and found that OAs increased an average of 16.5 and 4.8 percentage points compared with pixel-based and texture-based RF, respectively. A residual CNN (ResNet) and a recurrent neural network (RNN) were combined into an end-to-end architecture. It learned joint spectral–spatial–temporal feature representations based on multiseasonal Sentinel-2 imagery [21]. Results revealed the effectiveness of using temporal information for urban land cover classification. A deep CNN composed of residual learning and the Squeeze-and-Excitation block was introduced to classify LCZ as remote-sensing scene classification in a case study of metropolitan China [22]. However, these methods all used the image patches as inputs for classification. Although the local information within a patch is well mined, the global nonlocal information distributed in the whole image is not considered.

C. Graph Convolutional Network

GCN [23] has been a trending topic in recent years for its powerful learning ability on non-Euclidean data. It extracts features and transmits information between graph nodes to connect a regional and even global feature, which is not considered in CNN. Liu *et al.* and Wan *et al.* [24]–[26] adopted a superpixel-based approach to encoding an image to a graph. Pixels with similar texture features were first clustered into irregular pixel blocks, and graph nodes were extracted based on these superpixels. Each superpixel was connected to neighboring nodes that had a common edge. This method can not only describe texture features well but also reduce computational pressure. At the
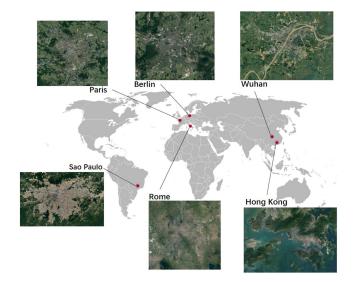

Fig. 2. Study areas with images from Google Earth.

same time, a graph node can not only connect with other nodes in the surrounding range. Hong *et al.* [27] proposed a method to establish global nodes relations, in which each pixel first found several pixels with the most similar spectral signature globally and then built connections. The article also discussed several fusion strategies of GCN and CNN and found that concatenation fusion can achieve better results.

These methods have shown their superiority in encoding nonlocal information to facilitate classification in the hyperspectral images. According to the characteristics of LCZ mapping, LCZ classification can also be analyzed and processed using these novel concepts and technologies.

III. STUDY AREA

A. Study Area

The research areas are six cities worldwide, namely, Rome, Hong Kong, Paris, Berlin, Sao Paulo, and Wuhan, as depicted in Fig. 2.

These six cities are spread over three continents. Geographical conditions in six cities vary considerably. Berlin is the capital and largest city of Germany. The city lies at the heart of the North German Plain, and the Spree River runs through the city center. Rome, the capital of Italy, is located in the central portion of the Italian peninsula. The city is monocentric. Thus, building density increases toward the center. Two great lakes, Lake Bracciano and Lake Albano are located in the northwest and southeast of the city. Wuhan is the capital city of Hubei Province in China. The city lies in the middle reaches of the Yangtze River, which flows through the center of the entire city, splitting the area in two. Thus, Wuhan has a highly developed water system and dots with thousands of lakes. Hong Kong is located in the south of China, east of the Pearl River estuary. The city is composed of three regions: 1) Hong Kong Island, 2) Kowloon Peninsula, and 3) New Territories. It has a high population density and land use density in the world. Sao Paulo is the largest and most prosperous city in Brazil. The city sits in a shallow basin with

TABLE I
DESCRIPTION OF CLIMATE IN SIX CITIES

City	Description of climate
Berlin	The transitional type between temperate maritime climate and temperate continental climate
Rome	Mediterranean climate with dry summers and cool, humid winters
Wuhan	Subtropical monsoon climates
Hong Kong	Humid subtropical climate with a hot and humid summer
Sao Paulo	Subtropical monsoon humid climate with mild, generally warm, and temperate summers
Paris	Temperate maritime climate

TABLE II
IMAGE ACQUISITION DATE FOR EACH CITY

City	Image acquisition date
Berlin	April 10, 2015
Rome	July 17, 2015
Wuhan	November 10, 2020
Hong Kong	October 18, 2015
Sao Paulo	August 3, 2014
Paris	September 27, 2015

TABLE III
NUMBER OF SAMPLES IN SIX CITIES

No.	class name	Berlin	Rome	Wuhan	Hong Kong	Sao Paulo	Paris
LCZ-1	Compact high-rise	-	-	55	631	955	56
LCZ-2	Compact mid-rise	1534	1551	122	179	134	2692
LCZ-3	Compact low-rise	-	104	210	326	5246	-
LCZ-4	open high-rise	577	-	237	673	482	366
LCZ-5	open mid-rise	2448	1495	339	126	244	426
LCZ-6	open low-rise	3901	480	129	120	1813	2121
LCZ-7	Lightweight low-rise	-	-	94	-	-	-
LCZ-8	Large low-rise	1483	435	251	137	1704	549
LCZ-9	Sparsely low-rise	632	-	8	-	258	-
LCZ-10	Heavy industry	-	51	424	219	179	-
LCZ-A	Dense tree	2868	284	607	1616	5758	2403
LCZ-B	Scattered trees	928	555	260	407	276	186
LCZ-C	Bush, scrub	605	-	65	691	-	-
LCZ-D	Low plants	2978	984	311	568	200	154
LCZ-E	Bare rock or paved	-	-	110	-	24	161
LCZ-F	Bare soil or sand	329	-	294	-	129	-
LCZ-G	Water	1417	500	2766	2579	2169	71
Total		19700	6439	6282	8272	19571	9185

low mountains to the west. Paris, the capital and largest city of the French Republic, is located in the middle of the Paris Basin, straddling the Seine River.

The Description of the climate of six cities is listed in Table I. Although Hong Kong, Wuhan, and Sao Paulo have subtropical monsoon climates, their geographical environments are quite different. In addition, the urban structures in the six cities vary, allowing us to demonstrate the robustness of the proposed approach.

B. Satellite Input and Reference Data

The Landsat-8 scene and reference data of five cities, namely, Berlin, Rome, Sao Paulo, Hong Kong, and, Paris were downloaded from the 2017 IEEE GRSS data fusion contest. It was hosted by the Image Analysis and Data Fusion Technical Committee in partnership with WUDAPT and Geo-Wiki. The acquisition date of the Landsat-8 scene is shown in Table II. We selected nine of the 11 bands in Landsat-8 as input data, namely, bands 1–7, 10, and 11, with a 100-m spatial resolution. Panchromatic and cirrus bands were removed from the feature. Because spectral information obtained by the panchromatic band overlaps greatly with other bands. Also, the cirrus band was mainly used for cirrus cloud detection instead of urban mapping. No additional extracted metrics or external datasets were utilized.

The Sentinel-2 scene of Wuhan was downloaded from the US Geological Survey Earth Explorer site,[1] then atmospheric-corrected, and clipped to cover the whole city. Raw reference data were downloaded from WUDAPT datasets. It was created by an urban expert as a Google Earth Keyhole Markup Language file [28]. Then, we visually checked and adjusted it using the

Google Earth desktop environment. After that, the adjusted reference data were projected and rasterized to align with the image.

The number of LCZ classes varied from city to city because of the different structural and architectural characteristics. There were 10 LCZ classes (6 built types and 4 land cover types) in Rome; 13 LCZ classes (8 built types and 5 land cover types) in Hong Kong; 12 LCZ classes (6 built types and 6 land cover types) in Berlin; 11 LCZ classes (6 built types and 5 land cover types) in Paris; and 16 LCZ classes (9 built types and 6 land cover types) in Sao Paulo; 17 LCZ classes (10 built types and 7 land cover types) in Wuhan. The sample size varied by category and by city as shown in Table III.

In Rome, Hong Kong, and Paris, 10%, 1%, and 89% of the samples per class were randomly selected for training, validation, and testing, respectively. Each sample represented a pixel. As for Wuhan, the ratio was 30%, 1%, and 69% because of its highly uneven sample distribution, especially for LCZ-G, accounting for nearly half of the total sample. The rates in Berlin and Sao Paulo were 5%, 1%, and 94% because they had a larger sample size. We adopted three commonly used indexes for accuracy assessment: 1) OA, 2) Kappa coefficient, and 3) average accuracy (AA). OA represents the ratio of the number of pixels correctly classified to the total number of pixels, representing the classification accuracy. The Kappa coefficient is a ratio. It represents the reduction in classification errors with the proposed network compared with the utterly random classification. The recall rate represents the ratio of the correctly classified samples to the actual samples in each category, and AA is the average of all the recall rates.

IV. PROPOSED METHODS

In this section, we first review the basic definition and notation of GCNs to provide some preliminary knowledge.

Then, we introduce the method we proposed, including the overall framework and three specific modules.

A. Review of GCN

Inspired by the remarkable job of convolution operation in Euclidean space, GCN was proposed to extract the features of non-Euclidean graph data. It continuously aggregates features from adjacent nodes. In our case, an undirected graph is employed, defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertexes, and $\mathcal{E}$ is the set of edges. The adjacency matrix A indicates whether an edge exists between two nodes. The edge describes the weights of nodes. The elements of A are computed as follows:

$$A_{i,j} = \begin{cases} 1, & if \ x_i \in Nei(x_j) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where vectors x_i and x_j are the spectral features of the nodes v_i and v_j, respectively, and $Nei(x_j)$ represents the set of neighbor nodes of x_j. In addition to dichotomy, A can be calculated using a dot-product similarity function [29]

$$A_{i,j} = \psi(x_i) \cdot \phi(x_j)^T \tag{2}$$

where $\psi(x_i) = W_\psi x_i$ and $\phi(x_j) = W_\phi x_j$ are two embeddings to embed the input into a linear vector space and calculate the similarity in the embedded space. Given A, the Laplacian matrix (L), and degree matric (D) are calculated as follows:

$$L = D - A \text{ and } D_{i,i} = \sum_j A_{i,j} \tag{3}$$

where D is a diagonal matrix that denotes the degree of A. And L is often symmetrically normalized as follows: $L_{\text{sym}} = I - D^{\frac{1}{2}} A D^{-\frac{1}{2}}$, where I is the identity matrix.

Because the number of neighbors of each vertex is different, graph convolution in the spatial domain has become a tough task. Hence, the graph convolution is defined in the frequency domain by the convolution theorem via the graph Fourier ($\mathcal{GF}$) transform. By spectral decomposition on L, $L = U \Lambda U^T$, we take U as the basis functions of $\mathcal{GF}$, and the convolution on a graph is formulated as follows:

$$\mathcal{G}[g*x] = U\left\{ [U^T x] \cdot [U^T g] \right\}. \tag{4}$$

Take $U^T g$ as g_θ, $g_\theta = \text{diag}(\theta)$ is a function of eigenvalues of L (i.e., $g_\theta(\Lambda)$). $g_\theta(\Lambda)$ is approximated by the K-order truncation of Chebyshev polynomials to reduce the computational complexity of eigendecomposition [30], and nodes are only influenced within their K-order neighbors. Finally, the graph convolution operation is simplified as (5). Only the first-order neighbors of nodes are considered, and some normalization technique is adopted [21]

$$X^{l+1} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X^l W^l \right) \tag{5}$$

where X^{l+1} and X^l are the output and input, respectively; $\sigma(\cdot)$ denotes the activation function, and W^l is the trainable parameter in the lth layer.

B. Proposed Network

The proposed method integrates residual network (ResNet), regional graph convolutional network (RGCN), and global graph convolutional network (GGCN) to form an end-to-end classification framework. The classification results are derived directly from the original images without additional cutting and feature extraction, as shown in Fig. 3.

First, we introduce 1×1 convolution as a pretreatment before graph convolution and ResNet to generate a robust and representative deep feature from original scenes. In Fig. 3, this module consists of two stacked layers. There is a batch normalization, a convolution layer, and an activation function in each layer. The convolution is formulated as follows:

$$X_{\text{pre}} = \sigma(W \cdot \text{BN}(X) + b) \tag{6}$$

where $X \in \mathbb{R}^{H*W*D}$ denotes the original image; H, W, and D denote the height, width, and band dimensions, respectively; $X_{\text{pre}} \in \mathbb{R}^{H*W*B}$ is the output of preprocessing; $\sigma(\cdot)$ denotes the activation function; $\text{BN}(X)$ is the batch-normalized X; W and b represent the weight with 1×1 kernel and bias, respectively. The operational symbol ($\cdot$) indicates the inner product. This module broadens the channels of the network and integrates the distinction of all bands. After that, more stable features are generated for the following process.

Next, three components are seamlessly integrated into an end-to-end trainable architecture to enable the discriminative local, regional, and global information to be combined for urban land cover classification. After that, the feature maps generated by three modules are stacked and fed into a Softmax classifier to calculate the probability of each category at each pixel. Finally, the cross-entropy loss function is utilized for training the whole network.

The whole network adopts the form of full convolution instead of the full connection layer to avoid an excessive number of parameters. We will explain the structure of the three modules in the following sections.

1) ResNet for Local Features: The ResNet module exploits the environmental context in the local area. Local building structure, surface materials, and vegetation interact to form a local microclimate, so the local feature is crucial for the LCZ classification. CNNs and their variants have achieved state-of-the-art performance in local feature extraction. In addition, ResNet [31] has been proven to outperform other methods in multiple data sets for its short-cut connection. It is more conducive to network convergence and easy to optimize.

The structure adopted in our work is shown in Fig. 3. This module consists of two residual blocks, each composed of a skip-connection and two convolution layers. The convolution operation multiplies the 3×3 pixels around the central pixel by the convolution kernel and adds up to obtain the feature point in the next layer. The receptive field will expand with the convolution. However, it is still limited to a small local range. This module can be written as follows:

$$X_{\text{Res}} = X_{\text{pre}} + \sigma(BN(W \cdot X_{\text{pre}} + b)) \tag{7}$$

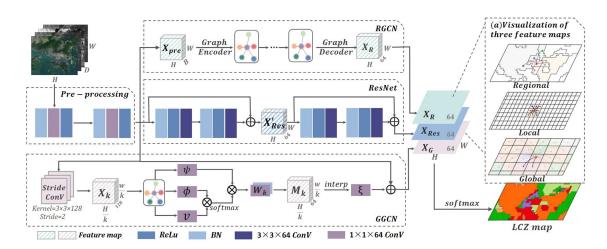where X_{Res} presents the output of the ResNet module.

Fig. 3.   Overall flowchart of the network architecture. (a) Visualization of features of three modules; the connections between nodes are also represented.
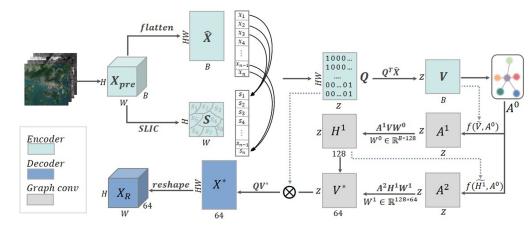

Fig. 4.   Graph encoder/decoder in the RGCN module.

2) Regional Graph Convolutional Network: In the LCZ classification, a large patch size helps to capture urban environmental features, and large surrounding information leads to a good performance [22]. However, in a fully convolutional network, each pixel can only obtain local information within a small range via ResNet, as in Fig. 3(a). Therefore, the RGCN network expands the receptive field to a larger range. The patch-based method [22] adjusts the size of the input patch to obtain a wide range of surroundings, but this information is aggregated indiscriminately. While in this article, we first cluster pixels based on their similarity and proximity to generate blocks and then aggregate the different blocks by weights to achieve a better regional interpretation.

First, the Simple Linear Iterative Clustering (SLIC) algorithm [32] is employed in our work to generate image segmentation (see Fig. 4). SLIC algorithm can generate compact and internally uniform superpixels, which is consistent with the definition of LCZ. $Z = \lceil \frac{(H \times W)}{\delta} \rceil$ is the number of superpixels. δ is the parameter controlling segmentation scale. $S = \{S_1, S_2, \ldots, S_Z\}$ indicates the superpixel set. Each superpixel is denoted as $S_i = \{x_j\}_{j=1}^{Ni}$, where x_i and N_i denote the pixels and number of pixels in the ith superpixel, respectively.

Second, an image-graph converter is utilized to create a regional-level graph. An index between regions and pixels is built as the converter. The index matrix $Q \in \mathbb{R}^{HW \times Z}$ is the one-hot encoding and can be presented as follows:

$$Q_{i,j} = \begin{cases} 1, & if\ \hat{X}_i \in S_j \\ 0, & otherwise \end{cases} \qquad (8)$$

where $\hat{X}$ is the flattened X, and $\hat{X}_i$ denotes the *i*th pixel in $\hat{X}$. Each row is a one-hot vector indicating which region the pixel belongs to, and the corresponding position is marked as one. Next, the features of each superpixel are extracted into a graph node. The encoder operation is computed by matrix multiplication

$$V = \hat{Q}^T\ \hat{X} \qquad (9)$$

where $\hat{Q}$ denotes the normalized Q by column to eliminate the negative impact of extreme values; the graph nodes $V = \{V_1, V_2, \ldots, V_Z\} \in \mathbb{R}^{Z \times D}$ can also be computed in elements form

$$V_i = \frac{1}{N_i} \sum_{j=1}^{Ni} x_j. \qquad (10)$$

That is, the feature of a superpixel is calculated from the mean of pixels within this segmentation. Hereto, an image has been encoded into a graph node. Thus, the adjacent matrix can be calculated, as in (1)

$$A_{i,j}^0 = \begin{cases} 1, & if \ V_i \in Nei\,(V_j) \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

The initial $A^0 \in \mathbb{R}^{Z \times Z}$ has only zero and one, where one exists only if two superpixels have a boundary, and the similarity between adjacent areas is difficult to reflect. Thus, we synchronously update A^0 to the range $[0, 1]$ to better describe the relation among nodes. We renew the adjacent matrix as follows:

$$A^{l+1} = \ f\left(\tilde{H}^l, A^0\right) \ = \ \sigma\left(\theta\left(\tilde{H}^l\right) \cdot \theta\left(\tilde{H}^l\right)^T\right) \odot A^0 + I \quad (12)$$

where H^l is the output of the graph convolutional in the l th layer, $\tilde{H}^l = BN(H^l)$; $\theta(\cdot)$ is a linear transformation to project features to the embedded space and calculates their similarity according to (2), and $\sigma(\cdot)$ denotes sigmoid activation, and operator $\odot$ indicates Hadamard product. With this update rule, the two layers of dynamic graph convolution are carried out according to (5).

Firstly, the region-level graph feature is reprojected to a pixelwise image by a graph decoder. As a reverse process to (9), the decoder is computed as follows:

$$X_R = \ \text{reshape}\,(QV^*) \quad (13)$$

where V^* denotes graph nodes of the output, and $\text{reshape}(\cdot)$ converts the feature to the same length and width as the original image.

In summary, the RGCN module first converts the image into a graph by an encoder. Next, the dynamic GCN is utilized to aggregate regional information. Finally, the graph is reprojected to the 3-D shape to concatenate with the CNN-processed feature.

3) Global Graph Convolutional Network: The same LCZ type may be located at a larger distance in the city. For example, Wuhan is a typical multicentric city with three city centers far away from each other. Compact low-rise class is distributed in different parts of Sao Paulo. RGCN cannot link these regions spaced far apart from each other. Hence, GGCN is proposed to build a global relationship between LCZs of the same type. It creates links for every two nodes and adjusts the weights of these links by calculating intraclass similarity. Thus, LCZs of the same type have larger weights to aggregate and extract more presentative features.

First, a downsampling operation is adopted to reduce the number of nodes and build a lightweight fully-connected graph

$$X_{\text{pre}} \in \mathbb{R}^{H*W*B} \Rightarrow X_k \in \mathbb{R}^{\frac{H*W}{k^2}*B} \quad (14)$$

where X is the downsampled feature, and k denotes the downsampling rate. We adopt a convolution downsampling method by three layers of continuous convolution, each with a stride of two and kernel size of 3×3, resulting in $k = 8$.

Next, each clustered pixel is treated as a graph node. Thus, the adjacency matrix $A_k \in \mathbb{R}^{\frac{H}{k}*\frac{W}{k}}$ is built to diffuse information

across nodes, and A_k is computed as follows according to (3)

$$A_k = \ \psi\,(X_k) \odot \ \phi(X_k)^T \quad (15)$$

where $\odot$ is a dot product operation, and $\psi(\cdot)$ and $\phi(\cdot)$ are two convolution operations with a kernel size of 1×1. Specifically, A_k measures the similarity between one node and all other nodes and changes in sync with the convolution process. The GGCN also follows a variation of (5) to reduce storage and computation

$$M_k = A_k \, v\,(X_k)\, W_k$$
$$= \psi\,(X_k) \odot \left(\phi(X_k)^T \odot v\,(X_k)\right) W_k \quad (16)$$

where $v(\cdot)$ is also a 1×1 convolution operation, and $v(X_k)$ works as the feature X^l in (5). Variable A_k is equal to $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ in (5).

After the GCN, a new feature M_k is reprojected to the initial space ($\mathbb{R}^{H*W*B}$) by upsampling. Different from the downsampling convolution, upsampling employs an interpolation operation to match the original size. Hence, the output feature is written as follows:

$$X_G = X_{\text{pre}} + \xi\,(\text{interp}\,(M_k)) \quad (17)$$

where X_G denotes the ultimate module output, $\text{interp}(\cdot)$ is the bilinear interpolation, and $\xi(\cdot)$ presents a 1×1 convolution to transform the output channel dimension.

In summary, the GGCN module builds a lightweight fully connected graph to adapt the feature through global information and provide complementary information for ResNet and RGCN.

After the process of three branches, we obtain three diverse features spanning different ranges, namely, local, regional, and global. The ultimate feature (X_u) is calculated as follows:

$$Y = \text{Softmax}\,(\text{Conv}\,(\text{concat}\,(X_R, \ X_{\text{Res}}, \ X_G))) \quad (18)$$

where $\text{concat}(\cdot)$ denotes the concatenation, $\text{Conv}(\cdot)$ is a convolution operation to adjust the dimensions for ensuring that the output is the number of classes to be predicted, and Y is the final output of the whole network.

C. Network Training

The detailed network structure and feature map size are described in Fig. 3. The configurations for hyperparameters of three modules, namely, ResNet, RGCN, and GGCN, are detailed as follows. In RGCN, the segmentation rate of SLIC was set to 100 to avoid excessive or incomplete fragmentation. In GGCN, the stride convolution downsampling parameter was set to k=8. Batch normalization [33] was adopted with the 0.9 momenta. Adam [34] was used to optimize the networks. The learning rate was set to 0.0005, and the epoch was 200 to ensure a well-trained network. The model with the lowest loss in the validation set was saved as the best model for the test set. All works were implemented with Python-3.7 and PyTorch-1.3.1. The experimental environment was a GTX-2080Ti GPU.
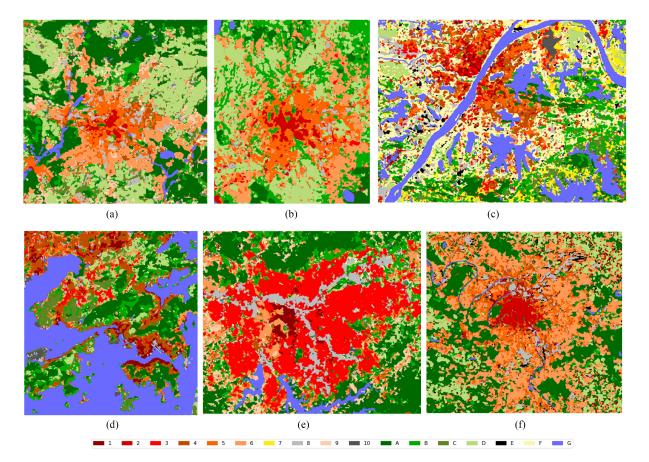
Fig. 5.   LCZ maps of the study cities. (a) Berlin. (b) Rome. (c) Wuhan. (d) Hong Kong. (e) Sao Paulo. (f) Paris.

V. RESULT

A. LCZ Maps of the Six Cities

Fig. 5 shows the developed LCZ maps, and the corresponding reference data are depicted in Fig. 6. In summary, the urban building types and natural types can be clearly distinguished. Water can also be accurately identified on maps. This work will detail the specific urban structure of each city in the following.

Berlin radiates outward in the shape of a hexagonal star. The core of this hexagonal star is LCZ-2 compact midrise, surrounded by a circle of LCZ-5 open midrise, and then extends out into suburbs as LCZ-6 open low-rise. The natural classes are dominated by LCZ-A type dense trees and LCZ-D low plants. The Havel River, running west of Berlin, can be clearly classified as LCZ-G. The Spree River, which flows through downtown, is not identified in the middle due to its narrow channel. However, the eastern stretch of the river can be distinguished.

As for Rome, it also appears monocentric center, and LCZ-2 dominates the urban center. Unlike Berlin, Rome has more intensive buildings (i.e., LCZ-2 and LCZ-3) dotted in the southwest and east of the city. The suburban areas are composed of open arrangements LCZ-5 and LCZ-6, far from the center.

The most common type in Wuhan is water LCZ-G, resulting from the Yangtze River and a host of lakes. Among the six cities, Wuhan has the most diverse urban types, LCZ 1-10, and the most complex urban structure. Compact buildings LCZ 1-3 are primarily distributed along the banks of the Yangtze River and its tributaries. The low/midrise industrial structure LCZ-10 is mainly located in a river bend close to the industrial port.

Hong Kong is surrounded by water LCZ-G and dense trees LCZ-A is the major natural type due to several Country Parks. Dense buildings, LCZ-1 and 4, are mainly observed in the northern part of Hong Kong Island and the southern part of the Kowloon Peninsula, opposite each other across the sea. The heavy industrial type LCZ-10 is mainly located near the Hong Kong International Airport.

In Sao Paulo, the primary type of architecture is LCZ-3 compact low-rise and is separated into chunks by the LCZ-8 large low-rise. Compact high-rise buildings LCZ-1 are concentrated only in the city center. The water LCZ-G in the picture presents two large reservoirs.

Paris is monocentric and has a regular circular center, principally LCZ-2 compact midrise. The city radiates suburbs mainly along the Seine River LCZ-G. The river is discernible and cuts through the city center. The bare rocks/paved areas LCZ-E is scattered in the suburbs. Suburban and natural feature types are mixed, with no clear boundaries.

In summary, the various urban structures and types of natural features can be well distinguished by our method. Water distribution can also be clearly observed. The resulting LCZ maps are visually satisfactory.
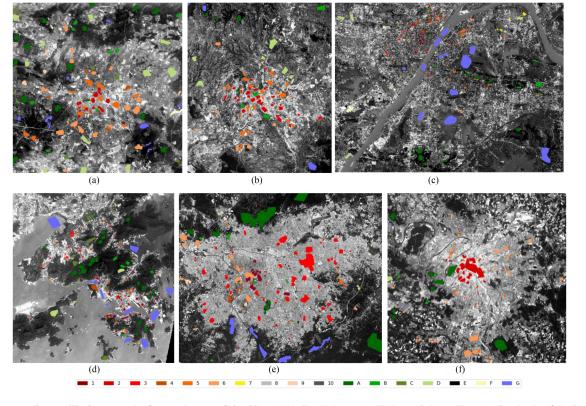
Fig. 6. Corresponding satellite images and reference datasets of six cities. (a) Berlin. (b) Rome. (c) Wuhan. (d) Hong Kong. (e) Sao Paulo. (f) Paris.

B. Quantitative Evaluation

The confusion matrix is adopted to display quantitative evaluation. In Fig. 7, the accuracy evaluation indexes are presented in the upper right of each diagram. The OA and Kappa of all cities all exceed 90%. The lowest OA=0.937 and Kappa=0.924 are observed in Hong Kong, and Wuhan has the highest OA=0.977 and Kappa=0.971. By contrast, AA is 5% lower than OA and Kappa on average. Because AA indicator places the same emphasis on each type and is easily affected by some types with small sample sizes that cannot obtain sufficient training. Thus, AA reaches a maximum in Berlin with 0.919 and a minimum in Sao Paulo with 0.769.

In Berlin and Paris, the majority of samples are well classified, especially for LCZ-A and LCZ-G. LCZ-G is identifiable for its unique spectral characteristics in many cases. But LCZ-A and LCZ-B are usually difficult to tell apart for spectral similarity [22]. However, the LCZ-A in our work achieves high accuracy in all cities.

In Rome, all classes have an accuracy over 0.9, except for the LCZ-10, for its training samples are too few to fully extract the characteristics of this type. There is also slight confusion between building types LCZ-2, LCZ-3, and LCZ-5. Similar to Rome, such a mix of building types is also inevitable in Wuhan. LCZ-9 in Wuhan also has the lowest accuracy due to insufficient training samples.

Hong Kong endures a more complex misclassification. The LCZ-2 is misclassified because the building distribution in Hong Kong is relatively chaotic. The midrise buildings in LCZ-2 mix with a few high-rise buildings. Thus, it is easy to be classified into LCZ-1 and LCZ-4. Besides, LCZ-B scattered trees, LCZ-C bush, and LCZ-D low plants are confused with each other because they are pretty similar in the spectral domain.

Sao Paulo achieves a relatively high OA even though the classification does not seem to work very well. The reason is the uneven distribution of samples in some categories. For example, LCZ-E has a classification accuracy of only 0.41 because there are only 22 pixels of this type, and 9 are correctly classified. This condition has little influence on the overall evaluation of OA but only leads to a poor AA.

In general, except for some confusion caused by insufficient samples and sample quality problems, all types can achieve a good classification under the combination of characteristics at different scales.

C. Comparison of the Different Classification Models

In this section, three methods for LCZ classification are compared with our method, including RF, large-scale mapping of LCZs with CNN (LSM-CNN) [20], and CNN scheme S5-CNN [9]. The experiment is carried out in six cities, and the final results are shown below. LSM-CNN and S5-CNN train 300 epochs, with a learning rate of 0.001 and a patch size of 32×32. There are 100 estimators in RF. The training and test samples used in the comparison method are the same as those in the proposed method. LCZ maps of comparison methods for two cities are shown in Fig. 8.
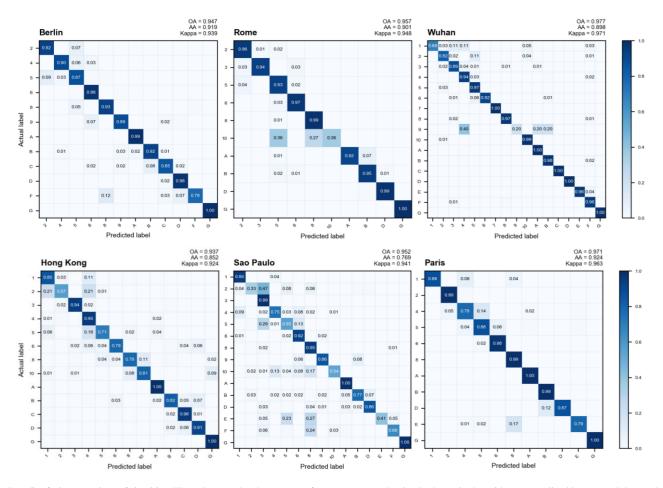
Fig. 7. Confusions matrixes of six cities. The color encodes the accuracy from zero to one. Decimals shown in the grid are normalized by row and do not show values less than 0.01.
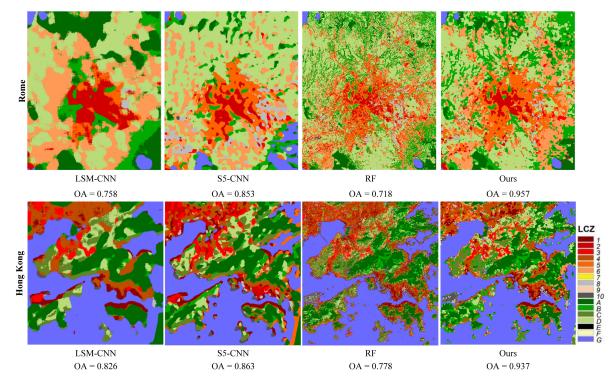

Fig. 8. LCZ maps of comparison methods for two cities.

TABLE IV
INDIVIDUAL CLASS, OA, AA, AND KAPPA OF ALL METHODS ON BERLIN AND ROME DATASETS

No.	Berlin				Rome			
	RF	LSM-CNN	S5-CNN	Ours	RF	LSM-CNN	S5-CNN	Ours
LCZ-2	0.79	0.86	0.88	**0.92**	0.77	0.96	0.88	**0.96**
LCZ-3	-	-	-	-	0.05	0.00	0.88	**0.94**
LCZ-4	0.24	0.33	0.61	**0.90**	-	-	-	-
LCZ-5	0.68	0.76	0.73	**0.87**	0.62	0.70	0.76	**0.93**
LCZ-6	0.91	**0.99**	0.90	0.98	0.56	0.29	0.86	**0.97**
LCZ-8	0.72	**0.96**	0.86	0.93	0.86	0.97	0.89	**0.99**
LCZ-9	0.51	0.26	0.72	**0.89**	-	-	-	-
LCZ-10	-	-	-	-	0.30	0.00	**0.61**	0.36
LCZ-A	0.97	0.96	0.96	**0.99**	0.95	0.80	0.88	**0.92**
LCZ-B	0.69	0.66	0.63	**0.92**	0.48	0.53	0.80	**0.95**
LCZ-C	0.48	0.65	0.84	**0.85**	-	-	-	-
LCZ-D	0.96	**1.00**	0.98	0.98	0.85	0.83	0.91	**0.99**
LCZ-F	0.62	0.19	0.51	**0.78**	-	-	-	-
LCZ-G	**1.00**	0.92	0.96	**1.00**	1.00	1.00	1.00	1.00
OA	0.834	0.860	0.862	**0.947**	0.718	0.758	0.853	**0.957**
AA	0.714	0.713	0.799	**0.919**	0.644	0.606	0.848	**0.901**
Kappa	0.801	0.841	0.841	**0.939**	0.656	0.698	0.823	**0.948**

The best metrics are emphasizes by boldface.

TABLE V
INDIVIDUAL CLASS, OA, AA, AND KAPPA OF ALL METHODS ON WUHAN AND HONG KONG DATASETS

No.	Wuhan				Hong Kong			
	RF	LSM-CNN	S5-CNN	Ours	RF	LSM-CNN	S5-CNN	Ours
LCZ-1	0.05	0.00	0.66	**0.68**	0.65	0.69	0.73	**0.85**
LCZ-2	0.11	0.49	0.81	**0.82**	0.28	0.01	0.66	**0.57**
LCZ-3	0.38	0.75	0.77	**0.89**	0.75	0.84	0.87	**0.94**
LCZ-4	0.29	0.73	0.87	**0.94**	0.60	0.81	0.77	**0.95**
LCZ-5	0.48	0.83	0.86	**0.97**	0.25	0.05	0.47	**0.71**
LCZ-6	0.07	0.65	0.88	**0.92**	0.32	0.05	0.35	**0.79**
LCZ-7	0.27	0.87	0.87	**1.00**	-	-	-	-
LCZ-8	0.18	0.93	0.95	**0.97**	0.24	0.41	0.57	**0.78**
LCZ-9	0.00	0.00	**0.80**	0.20	-	-	-	-
LCZ-10	0.67	**1.00**	0.96	0.99	0.53	0.82	**0.88**	0.81
LCZ-A	0.91	**1.00**	0.99	**1.00**	0.97	**1.00**	0.98	**1.00**
LCZ-B	0.28	0.72	0.91	**0.98**	0.48	0.36	0.59	**0.82**
LCZ-C	0.56	0.62	0.93	**1.00**	0.71	0.91	0.85	**0.96**
LCZ-D	0.81	**1.00**	**1.00**	**1.00**	0.57	0.73	0.75	**0.91**
LCZ-E	0.57	0.95	0.86	**0.96**	-	-	-	-
LCZ-F	0.44	0.88	0.85	**0.96**	-	-	-	-
LCZ-G	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
OA	0.736	0.918	0.947	**0.977**	0.778	0.826	0.863	**0.937**
AA	0.418	0.729	0.879	**0.898**	0.565	0.591	0.729	**0.852**
Kappa	0.653	0.892	0.931	**0.971**	0.735	0.791	0.837	**0.924**

The best metrics are emphasizes by boldface.

TABLE VI
INDIVIDUAL CLASS, OA, AA, AND KAPPA OF ALL METHODS ON SAO PAULO AND PARIS DATASETS

No.	Sao Paulo				Paris			
	RF	LSM-CNN	S5-CNN	Ours	RF	LSM-CNN	S5-CNN	Ours
LCZ-1	0.76	0.89	0.91	**0.95**	0.13	0.04	0.67	**0.88**
LCZ-2	0.02	0.00	0.10	**0.33**	0.94	**0.99**	**0.99**	**0.99**
LCZ-3	0.90	0.98	0.94	**0.99**	-	-	-	-
LCZ-4	0.36	0.34	0.38	**0.75**	0.26	0.58	0.67	**0.78**
LCZ-5	0.02	0.00	0.44	**0.55**	0.36	0.46	0.57	**0.88**
LCZ-6	0.75	0.88	0.80	**0.92**	0.91	**0.99**	0.95	0.98
LCZ-8	0.68	**0.97**	0.63	0.95	0.92	**1.00**	0.94	0.99
LCZ-9	0.41	0.00	0.83	**0.86**	-	-	-	-
LCZ-10	0.02	0.14	**0.73**	0.54	-	-	-	-
LCZ-A	0.98	0.99	0.97	**1.00**	0.98	**1.00**	0.99	**1.00**
LCZ-B	0.51	0.07	0.43	**0.77**	0.72	0.00	0.69	**0.99**
LCZ-D	0.46	0.06	0.64	**0.88**	0.82	0.00	0.23	**0.87**
LCZ-E	0.18	0.00	0.36	**0.41**	0.42	0.28	**0.83**	0.79
LCZ-F	0.11	0.01	0.37	**0.66**	-	-	-	-
LCZ-G	0.99	**1.00**	**1.00**	**1.00**	0.64	0.00	0.19	**1.00**
OA	0.828	0.886	0.869	**0.952**	0.862	0.901	0.927	**0.971**
AA	0.476	0.423	0.637	**0.769**	0.643	0.487	0.703	**0.924**
Kappa	0.786	0.856	0.837	**0.941**	0.821	0.871	0.905	**0.963**

The best metrics are emphasizes by boldface.

Tables IV–VI illustrate that our method significantly outperforms other methods and achieves the highest OA, Kappa, and AA in all the cities, followed by S5-CNN, LSM-CNN, and RF. This is due to the different acquisition and utilization of features in each method. CNNs can effectively extract the local context and thus outperform the pixel-based RF, which only considers the spectral characteristics of pixels without spatial characteristics. S5-CNN is superior to LSM-CNN because it uses a continuous convolutional layer to achieve a more stable representation, rather than connecting the pooling layer immediately after each convolutional layer. Furthermore, our method can generate regional and nonlocal features in addition to local representation to achieve the best accuracy.

However, there remain some exceptions. LSM-CNN surpasses S5-CNN in Sao Paulo. LSM-CNN improves the OA by sacrificing the accuracy of some categories with a small number of pixels. Thus, there are some small sample categories with accuracy as low as 0%. For example, due to highly uneven sample distribution in Sao Paulo, both OA and Kappa of the LSM-CNN result exceed S5-CNN, whereas S5-CNN still maintains a higher AA. In addition, S5-CNN has a higher classification accuracy for the categories with fewer samples and achieves the best OA in LCZ-9 (Wuhan) and LCZ-10 (Rome), exceeding our method. The main reason is that there are only 8 samples in LCZ-5 (Wuhan) and 51 samples in LCZ-10 (Rome).

VI. DISCUSSION

In this section, we design and conduct experiments from different perspectives to analyze the effectiveness of the proposed method.

A. Effectiveness of the Different Modules

In this network, different modules perform their respective roles and then combine to achieve the best results. Ablation

From Fig. 8, it is observed that the classification results of LSM-CNN and S5-CNN present oversmoothing results with large blocks. These methods that only use convolution to extract local features may be hard to distinguish built or land cover types with less spatial differences and similar spectral properties. Also, only considering the spectral differences between pixels, the results of RF show a fragmentized and granular effect that breaks the integrity within the zone. By fully utilizing the texture characteristics of the regional and the global features, the proposed method can obtain the appropriate classification scale without oversmoothing or fragmentized zones, which achieves a better classification result according to the real visual characteristics.

TABLE VII
OA (%) Indicator of the Different Module Combinations

Settings	Berlin	Rome	Wuhan	Hong Kong	Sao Paulo	Paris
CNN	92.6	93.9	96.7	92.4	94.5	94.8
CNN+RGCN	94.4	95.3	97.4	93.6	94.7	96.3
CNN+GGCN	93.1	95.3	97.3	93.3	94.9	96.2
Integrated	**94.7**	**95.7**	**97.7**	**93.7**	**95.2**	**97.1**

The best metrics are emphasizes by boldface.

experiments are conducted in this section to evaluate the effect of the different modules.

The CNN branch is considered a baseline and combined with the other two branches because local information is essential for LCZ mapping. On this basis, the GGCN and RGCN branches are added, and a network with three branches is integrated for the experiment. The accuracies of different module combinations are shown in Table VII. The results show that the network with RGCN or GGCN provides a significant improvement in OA compared with the performance of baseline (only CNN), which demonstrates the effectiveness of the GCN module. Moreover, the best performance is obtained by using all modules. The result shows the information extracted by GGCN and RGCN complement the information extracted by CNN, and the three promote each other.

B. Performance Under Limited Samples

In most cases, marking samples is usually costly in terms of human and time resources. Therefore, transferability among cities has been a hot topic discussed in the previous work. Previous work has attempted to train models on the training samples from one city and tests it on samples from other cities. However, the results suggest that the LCZ model cannot get satisfactory results by the regional shift [22]. The LCZ properties observed in arid desert environments [35] do not always match the standard value ranges [6], and city-to-city transfers are generally poor compared with the single city benchmark experiments [36]. Thus, small sample training could be the future trend, and it determines the applicability of a classification model.

To evaluate the proposed method under the condition of minimal training samples, we present the OA indices of different methods under various training sample sizes in Wuhan because it has a complete sample category. In Fig. 9, we randomly selected 5, 10, 15, 20, and 25 samples for each category as training samples in each experiment and two samples per class as verification samples in each experiment [26]. Each experiment was repeated 10 times to find the mean of the OA index. The OAs of the three deep-learning-based methods increase with the number of samples. Meanwhile, the RF is not sensitive to small sample size changes, and its accuracy presents a horizontal line. Our method is superior to all the comparison methods. Because the proposed method not only utilizes the local surroundings but also aggregates the information of the region and similar global nodes of each sample. These connections allow each sample to acquire more context information and a more robust feature
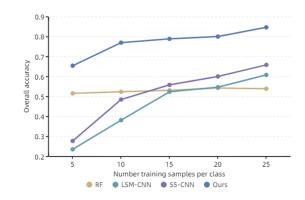

Fig. 9. Overall accuracy of different methods under different scales of training samples.

representation for different LCZ types. Therefore, high accuracy can be obtained by using small samples in our method.

VII. Conclusion

In this study, we develop a novel hybrid framework by combining two GCNs to assemble region- and global-level features to supplement the local-level features of the CNN in the LCZ classification and conduct LCZ mapping on six cities—Berlin, Rome, Wuhan, Hong Kong, Sao Paulo, and Paris—across the world. The weights of the graph edges update with the learned features dynamically. Furthermore, a graph encoder/decoder in RGCN and stride convolution in GGCN are adopted to decrease the computational cost and alleviate the structural incompatibility of GCN and CNN. Ablation tests indicate the complementarity between different modules to demonstrate the effectiveness of the unitized structure. This model also shows its validity in contrast experiment. We further explore the performance under different sizes of limited samples and find that our method exhibited the best small-sample learning ability among all the comparisons, which may serve as an alternative to transferability research and valuable guidance to tackle the insufficient-sample issue in LCZ mapping.

In the future, additional information, such as multitemporal and heterogeneous data, can be introduced to refine the model further. Long short-term memory and spatial–temporal attention mechanism can effectively analyze temporal features of multitemporal imagery. These features can work with the cross-regional features extracted by GCN to build a better understanding of the spatiotemporal dynamics and improve performance in the LCZ classification.

Acknowledgment

REFERENCES

[1] B. Cohen, "Urbanization, city growth, and the new United Nations development agenda role in achieving the least cost path to a sustainable," *Cornerstone Official J. Coal Ind.*, vol. 3, no. 2, pp. 1–72, 2015. [Online]. Available: www.cornerstonemag.net

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

[3] J. A. Fry *et al.*, "Completion of the 2006 National Land Cover Database for the conterminous United States," *Photogrammetric Eng. Remote Sens.*, vol. 77, no. 9, pp. 858–864, 2011.

[4] S. Kleeschulte and G. Büttner, "European land cover mapping—The CORINE experience," in *Proc. North Amer. L. Cover Summit*, 2006, pp. 31–44, doi: 10.11428/jhej1987.42.191.

[5] J. Chen *et al.*, "Global land cover mapping at 30 m resolution: A POK-based operational approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 7–27, 2015, doi: 10.1016/j.isprsjprs.2014.09.002.

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

[7] E. M. Ochola, E. Fakharizadehshirazi, A. O. Adimo, J. B. Mukundi, J. M. Wesonga, and S. Sodoudi, "Inter-local climate zone differentiation of land surface temperatures for management of urban heat in Nairobi City, Kenya," *Urban Clim.*, vol. 31, 2020, Art. no. 100540, doi: 10.1016/j.uclim.2019.100540.

[8] S. Ziaul and S. Pal, "Analyzing control of respiratory particulate matter on land surface temperature in local climatic zones of English Bazar Municipality and surroundings," *Urban Clim.*, vol. 24, pp. 35–50, 2018, doi: 10.1016/j.uclim.2018.01.006.

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

[10] B. Bechtel, L. See, G. Mills, and M. Foley, "Classification of local climate zones using SAR and multispectral data in an arid environment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3097–3105, Jul. 2016.

[11] C. C. Fonte, P. Lopes, L. See, and B. Bechtel, "Using OpenStreetMap (OSM) to enhance the classification of local climate zones in the framework of WUDAPT," *Urban Clim.,* vol. 28, 2019, Art. no. 100456, doi: 10.1016/j.uclim.2019.100456.

[12] N. Yokoya *et al.*, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, May 2018.

[13] G. Cheng, X. Xie, J. Han, L. Guo, and G. S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020.

[14] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[15] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.

[16] M. Liu, Q. Shi, A. Marinoni, D. He, X. Liu, and L. Zhang, "Super-resolution-based change detection network with stacked attention module for images with different resolutions," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–27, 2021, doi: 10.1109/TGRS.2021.3091758.

[17] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10348–10363, Dec. 2021.

[18] Q. Shi, M. Liu, S. Li, X. Liu, F. Wang, and L. Zhang, "A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, 2021, doi: 10.1109/TGRS.2021.3085870.

[19] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[20] J. Rosentreter, R. Hagensieker, and B. Waske, "Towards large-scale mapping of local climate zones using multitemporal Sentinel 2 data and convolutional neural networks," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111472, doi: 10.1016/j.rse.2019.111472.

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

[22] 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. Photogramm. Remote Sens.*, vol. 164, pp. 229–242, 2020, doi: 10.1016/j.isprsjprs.2020.04.008.

[23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Representations*, 2017, pp. 1–14.

[24] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2020.

[25] S. Wan, C. Gong, P. Zhong, S. Pan, G. Li, and J. Yang, "Hyperspectral image classification with context-aware dynamic graph convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 597–612, Jan. 2021.

[26] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657–8671, Oct. 2021.

[27] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[28] L. K. Ren and Y. Xu, "Wuhan WUDAPT level 0 training data is licensed under CC-BY-NC-SA 4.0 by the Chinese University of Hong Kong," 2016. [Online]. Available: www.wudapt.org/cities/raw-training-area-download/

[29] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7794–7803, 2018.

[30] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmonic Anal.*, vol. 30, no. 2, pp. 129–150, 2011.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.

[32] R. Achanta, A. Shaji, K. Smith, and A. Lucchi, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[34] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–9.

[35] C. Wang, A. Middel, S. W. Myint, S. Kaplan, A. J. Brazel, and J. Lukasczyk, "Assessing local climate zones in arid cities: The case of Phoenix, Arizona and Las Vegas, Nevada," *ISPRS J. Photogrammetry Remote Sens.*, vol. 141, pp. 59–71, 2018, doi: 10.1016/j.isprsjprs.2018.04.009.

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

Yedan Yu received the B.S. degree in sciences and techniques of remote sensing in 2019 from Wuhan University, Wuhan, China, where she is currently working toward the M.S. degree in photogrammetry and remote sensing.

Her research interests include graph convolution, land use classification, remote-sensing image processing, and deep learning.

Jie Li (Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011 and 2016, respectively.

He is currently an Associate Professor with the School of Geodesy and Geomatics, Wuhan University. His research interests include image quality improvement, image superresolution reconstruction, data fusion, remote-sensing image processing, sparse representation, and deep learning.

Qiangqiang Yuan (Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006 and 2012, respectively.

In 2012, he joined the School of Geodesy and Geomatics, Wuhan University, where he is currently a Professor. He has authored/coauthored more than 70 research papers, including more than 60 peer-reviewed articles in international journals such as the IEEE TRANSACTIONS IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His current research interests include image reconstruction, remote-sensing image processing and application, and data fusion.

Dr. Yuan was the recipient of the Youth Talent Support Program of China in 2019 and the Top-Ten Academic Star of Wuhan University in 2011 and was also the recipient of the 2014 Hong Kong Scholar Award from the Society of Hong Kong Scholars and the China National Postdoctoral Council. He is an Associate Editor for IEEE ACCESS journal and has frequently served as a referee for more than 40 international journals for remote sensing and image processing.

Qian Shi (Senior Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, in 2015.

She is currently an Associate Professor with the School of Geography and Planning, Sun Yat-sen University, Guangzhou, China. Her research interests include remote-sensing image classification, including deep learning, active learning, and transfer learning.

Huanfeng Shen (Senior Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he joined the School of Resource and Environmental Sciences (SRES), Wuhan University, where he is currently a Luojia Distinguished Professor and an Associate Dean of SRES. He was or is the PI of two projects supported by the National Key Research and Development Program of China, and six projects supported by the National Natural Science Foundation of China. He has authored more than 100 research papers in peer-reviewed international journals. His research interests include remote-sensing image processing, multisource data fusion, and intelligent environmental sensing.

Dr. Shen is a Council Member of China Association of Remote Sensing Application, an Education Committee Member of Chinese Society for Geodesy Photogrammetry and Cartography, and a Theory Committee Member of Chinese Society for Geospatial Information Society. He is currently a member of the Editorial Board of the *Journal of Applied Remote Sensing* and *Geography and Geo-Information Science*.

Liangpei Zhang (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently a Chair Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He was a Principal Scientist for the China State Key Basic Research Project, from 2011 to 2016, appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has authored/coauthored more than 700 research papers and five books. He is the Institute for Scientific Information (ISI) highly cited author. He is the holder of 30 patents. He is also an Associate Editor or Editor of more than ten international journals. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology. He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS), and the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) student paper contest in recent years. He is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is the founding Chair of IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter.