A CO-TRAINING APPROACH TO THE CLASSIFICATION OF LOCAL CLIMATE ZONES WITH MULTI-SOURCE DATA

Yong XU[1], Fan MA[2], Deyu MENG[2], Chao REN[1], Yee LEUNG[1]

[1] Institute of Future Cities, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
[2] Institute for Information and System Sciences and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an, China

ABSTRACT

Local climate zone (LCZ) classification system provides standard urban morphological classification for urban heat island studies and weather and climate modelling. Based on the definition of the LCZ, various semi-supervised classification approaches have been proposed to generate LCZ maps for different cities using available satellite data. Given that the acquisition of training data is labor intensive, it is practical to develop new models that are suitable for LCZ classification for any cities without the need for training data/samples. In this study, a novel domain-adaptation co-training approach with self-paced learning is designed to generate LCZ maps for new cities with which valid training samples from existing cities are explored and transferred to new target cities for classification. Experimental results show that the proposed approach could derive LCZ maps for the four testing cities, with an overall accuracy of 69.8%, which is over 10% more accurate than conventional approaches. Compared with conventional approaches, the novel approach does not need prior knowledge about the target cities, and it can automatically generate worldwide LCZ maps to support urban-climate studies for cities in the world.

Index Terms— Local climate zone, self-paced learning, co-training.

1. INTRODUCTION

The concept of local climate zone (LCZ) was developed by Stewart and Oke (2012). According to the standard definition of a LCZ in terms of land surface cover, urban structure, material and human activity in urban areas, a city can be classified into 17 standard classes, and conditions in different LCZ classes can impact on urban heat island intensity and distribution pattern [1]. Compared with the conventional rural-urban classification system, the new LCZ classification system can help to explore the relationship between urban morphological types and their effect on urban heat island effects instead of urban-rural temperature difference. The LCZ classification has become an international standard system for urban heat island studies [2-10].

According to the availability of urban data, two methods have been developed to generate urban LCZ maps, namely, the geographic information system (GIS)-based method and remote sensing-based method [10]. The GIS-based method uses the actual urban GIS data (e.g., 3D buildings and land use/cover) to generate highly accurate LCZ maps. This kind of method first uses actual urban GIS data to generate a series of required parameters. Then, the local urban area with different parameters could be classified into different LCZ classes by referring to the standard definition of the parameters for each LCZ class. Based on the actual GIS data, Gál et al. (2014) generated LCZ maps for Szeged, Hungary. By combining both urban GIS data with open-source data like Google Maps and population maps, Puliafito et al. (2013) generated LCZ maps for Mendoza, Argentina.

The remote sensing-based approach is another popular method used to generate LCZ maps, given that the updated urban structure data may not be available for some cities. Based on the freely available Landsat satellite data, a new initiative called the World Urban Database and Access Portal Tools (WUDAPT; http://www.wudapt.org) has been designed to generate LCZ maps for cities worldwide [2, 3]. It aims to use free Landsat images to speedily develop LCZ data for various applications, like urban climate studies and air pollution analysis. Since different cities might have different urban structures and morphologies, Danylo et al. (2016) applied the conventional WUDAPT method to two Ukrainian cities with very different urban environments. To improve the results for some high-density Asian cities, Xu et al. (2017) exploited the use of textural features from multi-source satellite data to achieve better LCZ mapping results. Furthermore, Casonne et al. (2016) compared the popular object-based image analysis (OBIA) approach with the conventional pixel-based approach in generating LCZ maps and found that the accuracies of the object-based and pixel-based approaches were comparable but the inclusion of additional attributes slightly improved the object-based classification. Incidentally, Tuia et al. (2016) investigate the use of spatial interaction information under a random field framework to achieve better LCZ mapping results.

However, the acquisition of training data for the conventional LCZ mapping approaches is labor intensive, and developing new models without training data suitable for new cities is still a challenging issue [7, 8]. The availability of the Earth Observation data from various sensors (e.g., Landsat, Sentinel, and ASTER data) might make it possible to solve the transferability issue of training data, as it might be possible to obtain more generic features like cloud free, homogeneous and multi-temporal features from all acquisitions, which will simplify the WUDAPT workflow in a more efficient way [7, 8].

Thus, in this study, we investigate the use of training data from domain cities to map LCZ for new target cites and develop a novel domain-adaptation method to generate highly accurate LCZ mapping results with multi-source open data.

2. PROPOSED METHOD

Based on multi-source data, a novel co-training classification approach coupled with an advanced self-paced learning strategy is designed to generate LCZ maps for this contest. The proposed approach has several advantages, including in particular the new features [17] and the innovative method [18-22] recently developed by our team, and multi-models are well integrated to achieve an improved LCZ classification. As shown in Fig. 1, the proposed approach comprises four main stages. Details of each stage are explained below.

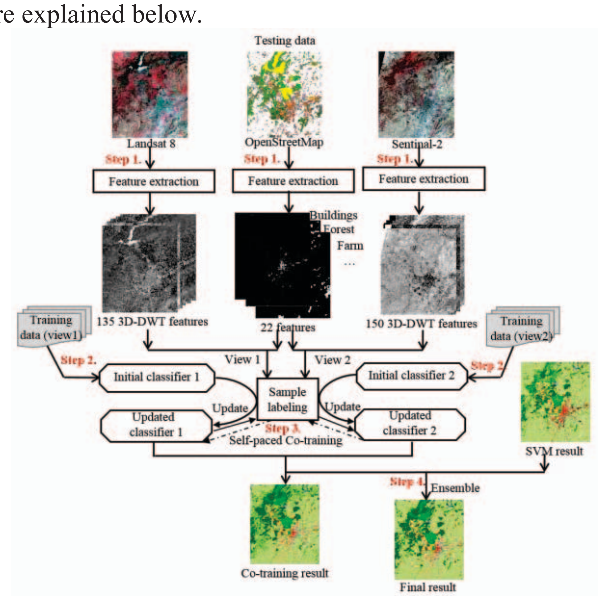

Fig. 1. Procedure of the proposed approach

2.1 Feature extraction

The three-dimensional discrete wavelet transform [17] is employed to generate spectral-spatial features from the original Landsat-8 and Sentinel-2 satellite data. In particular, 135 and 150 spectral-spatial features are extracted from the Landsat-8 and Sentinal-2 satellite data, respectively. Given that the OpenStreetMap (OSM) has 22 different land cover categories, 22 features representing proportions of different land covers are generated from the original OSM data. Thus, a total of 307 features are obtained from the three datasets.

2.2 Training classifiers

The well-known extreme gradient boosting (XGBoost) model is used to train two independent XGBoost classifiers with the provided training samples from five cities. Extended on previous work [14], the 135 spectral-spatial features derived from the Landsat data and the 22 land cover features obtained from the OSM data are combined to train the XGBoost classifier 1, while the 150 spectral-spatial features derived from the Sentinel-2 data and the 22 land cover features obtained from the OSM data are used to train the XGBoost classifier 2.

2.3 Co-training for domain adaptation and classification

The co-training approach with self-paced learning strategy is designed to adapt the model to suit the target cities [22]. For the target cities, the built XGBoost classifiers are used to generate two independent LCZ classification maps with probability estimations. Then, based on both results, high-confidence classified samples from classifier 1 are labelled and used to improve classifier 2, while the high-confidence classified samples from classifier 2 are used to improve classifier 1. By adding valid samples from target cities using the proposed co-training approach, the original classifiers (1 and 2) are iteratively modified, and the new classifiers are proven to be more suitable for the target cites with higher overall accuracy. Results of the co-training approach are the combination of the results from both classifiers.

2.4 Ensemble of multi-classifiers

To further improve our results, we investigate the combination of the co-training approach with other classifiers, such as support vector machine (SVM) and multi-layer perceptron (MLP), to enhance the final classification accuracy with the complementary information coming from different classifiers. Experimental results show that the combination of the SVM classifier and the proposed co-training approach further improves the accuracy by around 2-3%.

3. EXPERIMENTAL RESULTS

In this section, we first describe the datasets and implementation details and then present the results and performance analysis of the proposed approach.

3.1 Datasets

As shown in Fig. 2, the Image Analysis and Data Fusion Technical Committee (DFC) of the IEEE Geoscience and Remote Sensing Society (GRSS) (grss_dfc_2017 [7]) provides the datasets for nine cities, namely the datasets for Berlin, Hong Kong, Paris, Rome and Sao Paulo, as training data, while datasets for Chicago, Amsterdam, Madrid, and Xi'an are used for testing. For each city, three kinds of data are provided, including the Landsat-8, Sentinel-2 satellite, and OpenStreetMap (OSM) data.

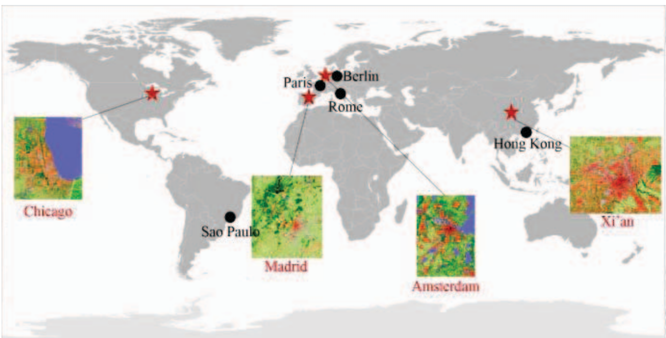

Fig. 2. Datasets from nine cities provided by 2017 IEEE GRSS DFC [7].

To validate the performance of the proposed approach, the datasets from five training cities with label data are used to conduct this study, in which Berlin, Hong Kong, Rome and Sao Paulo are used as training data, and Paris is used as the validation data.

3.2 Results and performance analysis

Using the proposed approach, Fig. 3(d) shows the LCZ mapping result for Paris, the original image is given in Fig. 3(a). For comparison, the results obtained by several other methods are also depicted in Figs. 3(b)-(c), with Fig. 3(b) showing the LCZ mapping result using the original XGBoost classifier, and Fig. 3 (c) showing the LCZ mapping result using the SVM classifier.

Two accuracy assessment indices, namely the overall accuracy (OA) and the kappa coefficient, are used to assess the results by comparing them with the actual validation data. Table 1 shows the accuracy statistics for all results of Paris, from which two findings can be summarized. First, the proposed approach generally obtains the best result, as the overall accuracy of the proposed approach (73.2%) is much better than those of the conventional methods like MLP, SVM, and XGBoost, with overall accuracies of 58.8%, 59.6%, and 58.4%, respectively. Second, the ensemble of different classifiers can improve the final generated results, as the result with the ensemble of SVM and self-paced co-training is 73.2%, which is much better than the result using either SVM or the self-paced co-training classifier.

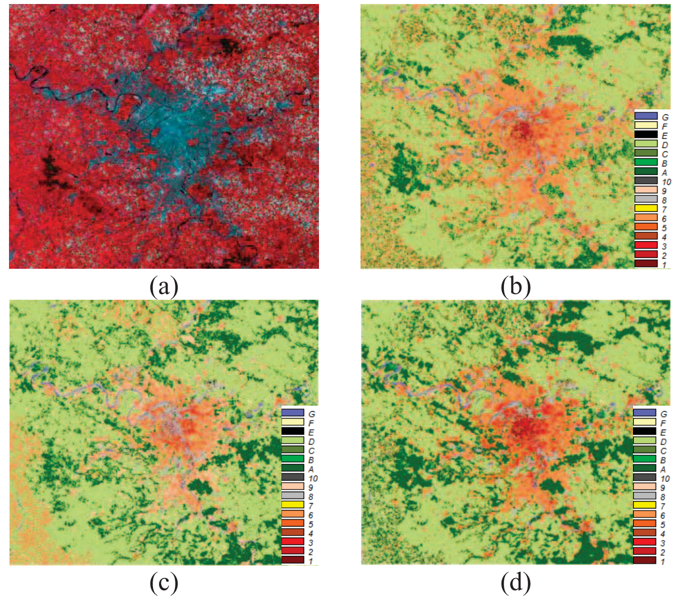

Fig. 3. LCZ mapping results for Paris using multi-source data. (a) Original satellite data; (b) LCZ mapping result using the XGBoost classifier; (c) LCZ mapping result using the SVM classifier; (d) LCZ mapping result using the proposed approach.

Table 1. Accuracy statistics for the results of Paris with different methods

Methods	MLP	SVM	XGBoost	SP_Cotrain	Proposed
OA (%)	58.8	59.6	58.4	**69.1**	**73.2**
Kappa	0.47	0.48	0.47	**0.61**	**0.66**

Moreover, the proposed approach has been successfully applied to generate high-quality LCZ mapping results for the four testing cities, namely Chicago, Amsterdam, Madrid, and Xi'an. Based on the proposed method, the overall accuracy of the results for the four testing cities is 69.8%, which is much better than the result obtained using the conventional XGBoost classifier (59.5%). Given that the proposed method does not require any prior knowledge about the targeted cities, this result is quite impressive and can certainly serve as a basis for further development.

However, the overall mapping accuracy is not satisfactory for some minor classes, like LCZ 9 and LCZ C, which might be due to the unbalanced data problem that is not well considered in our design. Our approach is based on a co-training idea by which high-confidence classified samples from target cities are iteratively selected as training data. However, these selected samples may still contain noise and tend to be dominated by some major classes, like LCZ A and LCZ D. Therefore, after several iterations, the improvement might become quite significant for some major classes (e.g., LCZ A), while some minor LCZ classes may be incorrectly classified. The unbalanced data problem requires further study.

4. CONCLUSIONS

A novel co-training-based approach is designed to generate high-quality LCZ maps for new target cities in this contest. Multi-source satellite data are first utilized to generate new spatial-spectral features by the three-dimensional discrete wavelet transform, a novel co-training classification method is then employed to generate high-quality LCZ maps. The proposed approach has several advantages: it makes use of new datasets, new features and the new domain-adaptation approach to achieve good performance. To further improve its performance, the ensemble of multi-classifiers is also applied and evaluated. Extensive experiments indicate that the proposed approach achieves much better LCZ mapping results than conventional approaches.

The method is part of the WUDAPT initiative, an international effort seeking to improve the acquisition, storage, and dissemination of large-scale urban morphological data (www.wudapt.org). The proposed approach can speed up data collection in the urban morphological mapping process, and reduce the manpower needed in climate-sensitive urban planning. The innovative and accurate method enables the quick assembly of city-related data needed by policy makers, city planners, and urban climate scientists to build smarter, healthier, and greener cities.

Acknowledgement

This research was supported by the VC's discretionary fund of The Chinese University of Hong Kong, the China NSFC project under contract 61373114, 61661166011, and 11690011, Macau Science and Technology Development Funds with No. 003/2016/AFJ, and 973 Program of China with No. 2013CB329404.

5. REFERENCES

[1] I. D. Stewart and T. R. Oke, "Local Climate Zones for Urban Temperature Studies," *Bull. Am. Meteorol. Soc*, vol. 93, no. 12, pp. 1879–1900, Dec. 2012.

[2] G. Mills, J. Ching, L. See, and B. Betchel, "An introduction to the WUDAPT project," presented at *the 9th International Conference on Urban Climat*e (ICUC9), Toulouse, France, 2015.

[3] 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, Feb. 2015.

[4] J. Ching, "A perspective on urban canopy layer modeling for weather, climate and air quality applications," *Urban Climate*, vol. 3, pp. 13-39, May. 2013.

[5] J. Feddema, G. Mills, and J. Ching, "Demonstrating the Added Value of WUDAPT for Urban Climate Modelling," presented at *the 9th International Conference on Urban Climate (ICUC9)*, Toulouse, France, 2015.

[6] C. Ren et al., "Local Climate Zone (LCZ) Classification Using the World Urban Database and Access Portal Tools (WUDAPT) Method: A Case Study in Wuhan and Hangzhou," presented at *the Fourth International Conference on Countermeasure to Urban Heat Islands (4th IC2UHI)*, Singapore, 2016.

[7] D. Tuia, G. Moser, B. Le Saux, B. Bechtel, and L. See, "IEEE GRSS Data Fusion Contest: open data for global multimodal land use classification," *IEEE Geosci. Remote Sens. Mag*, 2017.

[8] B. Bechtel, M. Demuzere, Y. Xu, M.L. Verdonck, P. Lopes, L. See, C. Ren, F. Van Coillie, D. Tuia, C.C. Fonte, A. Cassone, N. Kaloustian, O. Conrad, M. Tamminga, and G. Mills, "Beyond the urban mask: Local climate zones as a generic descriptor of urban areas – Potential and recent developments," *Jt. Urban Remote Sens*. Event JURSE, 2017.

[9] B. Bechtel and C. Daneke, "Classification of Local Climate Zones Based on Multiple Earth Observation Data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*, vol. 5, no. 4, pp. 1191–1202, Aug. 2012.

[10] B. Bechtel, L. See, G. Mills, & M. Foley, "Classification of Local Climate Zones Using SAR and Multispectral Data in an Arid Environment," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no.7, pp.3097-3105, Mar. 2016.

[11] T. Gál, B. Bechtel, & J. Unger, "Comparison of two different Local Climate Zone mapping methods," *presented at the 9th International Conference on Urban Climat*e (ICUC9), Toulouse, France, 2015.

[12] S. E. Puliafito, F. R. Bochaca, D. G. Allende, & R. Fernandez, "Green areas and microscale thermal comfort in arid environments: A case study in Mendoza, Argentina," *Atmospheric and Climate Sciences*, vol.3, no.3, pp.372-384. 2013.

[13] O. Danylo, L. See, B. Bechtel, D. Schepaschenko, and S. Fritz, "Contributing to WUDAPT: A Local Climate Zone Classification of Two Cities in Ukraine," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1841-1853, May. 2016.

[14] Y. Xu, C. Ren, M. Cai, Y.Y. Ng, & T. Wu, "Classification of Local Climate Zones Using ASTER and Landsat Data for High-Density Cities," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* May. 2017.

[15] A. Casonne, "Deriving Local Climate Zones from Remote Sensing data," *Master thesis*, Université de Strasbourg, 2016.

[16] D. Tuia, G. Moser, M. Wurm, and H. Taubenböck, "Land use modelling in North Rhine-Westphalia with interaction and scaling laws," *Jt. Urban Remote Sens*. Event JURSE, 2017.

[17] X. Cao, L. Xu, D. Meng, Q. Zhao, & Z. Xu, "Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification," *Neurocomputing*, vol.226, pp. 90-100. 2017.

[18] D. Zhang, D. Meng, & J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE transactions on pattern analysis and machine intelligence*, vol.39.no.5, pp.865-878. 2016.

[19] L. Jiang, D. Meng, S. I. Yu, Z. Lan, S. Shan, & A. Hauptmann, "Self-paced learning with diversity," *In Advances in Neural Information Processing Systems*, pp. 2078-2086. 2014.

[20] D. Meng, Q. Zhao, L. Jiang, "A Theoretical Understanding of Self-paced Learning," *arXiv*:1511.06049, 2015.

[21] Q. Zhao, D. Meng, L Jiang, Q. Xie, Z. Xu, A. Hauptmann, "Self-paced Matrix Factorization," *AAAI*, 2015.

[22] F. Ma and D. Meng, "Self-paced Co-training," *International Conference on Machine Learning (ICML)*, 2017. (Accepted)