

FUSING MULTI-SEASONAL SENTINEL-2 IMAGES WITH RESIDUAL CONVOLUTIONAL NEURAL NETWORKS FOR LOCAL CLIMATE ZONE-DERIVED URBAN LAND COVER CLASSIFICATION

Chunping Qiu¹, Michael Schmitt¹, Xiao Xiang Zhu^{1,2}

¹Signal Processing in Earth Observation, Technical University of Munich (TUM), Munich, Germany

²Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

ABSTRACT

This paper proposes a framework to fuse multi-seasonal Sentinel-2 images, with application on LCZ-derived urban land cover classification. Cross-validation over a seven-city study area in central Europe demonstrates its consistently better performance over several previous approaches, with the same experimental setup. Based on our previous work, we can conclude that decision-level fusion is better than feature-level fusion for similar tasks at similar scale with multi-seasonal Sentinel-2 images. With the framework, urban land cover maps of several cities are produced. The visualization of two exemplary areas shows urban structures that are consistent with existing datasets. This framework can be also generally beneficial for other types of urban mapping.

Index Terms— Sentinel-2, Classification, residual convolutional neural network (ResNet), urban land cover, long short-term memory (LSTM)

1. INTRODUCTION

The local climate zone (LCZ) scheme was originally proposed for climate relevant studies such as urban heat islands (UHIs), consistent across the world [1, 2, 3, 4]. LCZ-derived urban land cover is proposed for a hierarchical (from coarse to fine) local climate zone (LCZ) mapping as well as a detailed urban built-up area or human settlement mapping [5]. The LCZ-derived urban land cover scheme includes six classes: Compact built-up area, Open built-up area, Sparsely built, Large Low-rise/Heavy industry, Vegetation, and Water, by combining similar LCZ classes. It, on the one hand, includes all the land cover cases in a consistent manner across the world, thus is able to represent urban environments. Additionally, it can be further extended into the complete LCZ scheme.

Previously, we have mainly investigated Sentinel-2 images for this urban land cover classification task. This is because we are aiming at large-scale mapping with openly available data in the end. In addition, the five-day revisit time (at the equator) of Sentinel-2 makes it unprecedentedly feasible for further change detection and long term monitoring of

the urban extent [6] world wide. Specifically, we investigated multi-seasonal Sentinel-2 images, which demonstrates better results over the single-seasonal input, by using a ResNet architecture as the baseline network [5]. In this paper, we want to further exploit the potential of the multi-seasonal Sentinel-2 images for urban land cover classification, providing more methodological insights into this topic. Additionally, based on the best achieved mapping results, an analysis of the remaining challenges and possible solutions will be carried out. Based on the discussion, some future steps are designed, aiming more accurate urban land cover mapping.

2. DIFFERENT APPROACHES FOR FUSING MULTI-SEASONAL SENTINEL-2 IMAGES

The fusion of multi-seasonal Sentinel-2 images can be carried out on three different levels: observation-level, feature-level, or decision-level [7]. In previous work, we investigated the first two levels as follows:

- observation-level (st-ResNet): First, we exploited a classic ResNet architecture with four seasonal images stacked together as the input.
- feature-level (Re-ResNet): In addition, we exploited a four-stream ResNet, followed by a long short-term memory (LSTM) network, with the four seasonal images as the input to the four-stream ResNet, respectively.

We have shown that the feature-level fusion is better than the observation-level one, when the application is to map the urban land cover [5]. One step further, in this paper we propose a novel framework for decision-level fusion, which is illustrated in Fig. 1. In short, we exploit a four-stream ResNet to learn the spectral-spatial features from the multi-seasonal Sentinel-2 images and predict the urban land cover labels, and seamlessly integrate them into one end-to-end architecture by simply averaging the predicted class probability. In addition, class probability can also be predicted by the first three residual blocks, instead of four. Depending on the number of the predictions to be averaged, the framework is referred as

Res_ensemble_8 or Res_ensemble_4 in this paper. By doing so, the different decisions made independently from different seasons can be better harnessed for the urban land cover classification.

3. EXPERIMENTAL RESULTS

3.1. Study Areas and Datasets

In order to compare different fusion approaches in a fair manner, the experiments follow the same practice as used in our previous work [5]. For completeness of understanding, the study area, the datasets and the experimental setup are briefly described below. The study area consists of seven cities across Europe: Amsterdam, Berlin, Cologne, London, Milan, Munich and Paris. The (mostly) cloud free multi-seasonal Sentinel-2 imagery is processed with Google Earth Engine (GEE) [8] and the 10 meter and 20 meter bands are used. The reference ground truth data is from the LCZ42 dataset [9], and is further prepared by class combination and data augmentation as in [5]. Also, the accuracy assessment is carried out on absolutely balanced samples. Therefore, only two measures, overall accuracy (OA) and kappa coefficient, are used for accuracy analysis.

3.2. Comparative classification accuracy

The comparative classification accuracy achieved using different methods is shown in Tab. 1, where the first 7 rows are from our previous work [5], and the first four rows is to exploit the ResNet with single seasonal images as the input, and av-ResNet is to first independently exploit the ResNet for all four seasons, followed by averaging the accuracy of four seasons. An accuracy improvement can be achieved from the proposed fusion approach, compared to the previous efforts. Additionally, Res_ensemble_8 is better than Res_ensemble_4, for both OA and Kappa.

A further comparison of these approaches for all the seven test cases is shown in Fig. 2. From Fig. 2, it can be seen that above observations are consistent for most of the test cases.

Fig. 3 depicts the misclassifications, by combing the confusion matrices of all 7 test cases together. While most classes are classified correctly, class 3 (Sparsely built) tends to be classified into class 2 (Open built-up area) and class 5 (Vegetation).

3.3. Produced urban land cover maps

Urban land cover maps can be produced for all the test cities, using the trained Res_ensemble_8. The land cover map of the Munich city is shown in Fig.4 .

An zoomed-in view of two small areas of Munich, Germany is show in Fig. 5. The corresponding optical image from Google Earth and the Global Human Settlement (GHS)

Table 1: Comparative classification accuracy from different approaches. The measures are averaged over the 7 test cases. The results except are from our previous work [5], Res_ensemble_4 and Res_ensemble_8. The bold values represent the best accuracy achieved.

temporal information	approach	OA	Kappa
not considered	spring	82.7%	0.79
	summer	81.2%	0.77
	autumn	82.7%	0.79
	winter	77.9%	0.74
	av-ResNet	81.1%	0.77
considered	st-ResNet	79.8%	0.76
	Re-ResNet	84.0%	0.81
	Res_ensemble_4	85.3%	0.82
	Res_ensemble_8	86.7%	0.84

built-up grid [10] are also provided for comparison. Consistency between the produced land cover maps and the GHS dataset can be generally observed.

4. DISCUSSION

Both the accuracy measures and the processed land cover maps in Section 3 demonstrate better performance of the proposed framework, compared to several other approaches. It is suggested that decision-level fusion is better when fusing multi-seasonal Sentinel-2 images for similar urban land cover classification. When taking all the test cases into account, the OA and Kappa can be improved from 84.0% to 86.7%, and 0.81 to 0.84, respectively, as shown in Tab. 1, compared to a sophisticated Re-ResNet which integrates a four-stream ResNet and a LSTM. As explained in Section 2, the Re-ResNet falls into the feature-level fusion category, while the Res_ensemble_4 and Res_ensemble_8 are decision-level fusion. Why is decision-level fusion better for the task? The first possible reason is that the time steps (four seasons) are not enough. In addition, this is probably because of the employed multi-seasonal Sentinel-2 images processed using GEE, as described in [11]. When the cloud removal is carried out in GEE, we aggregated all the images in one season into one mostly cloud free image, which is already an equivalent observation-level fusion of the all available Sentinel-2 images within that season. As a result, the temporal resolution is not fully preserved in the input data, which might hinder the temporal information extraction when fusing multi-seasonal images. On the other hand, a fusion on the decision-level, which is equivalent to an ensemble of multiple classifiers, is able to robustly exploit the joint power of the multi-seasonal images, and better overcome other image distortion, even though the temporal information is not explicitly extracted.

While the produced maps are promising, as shown in Fig.

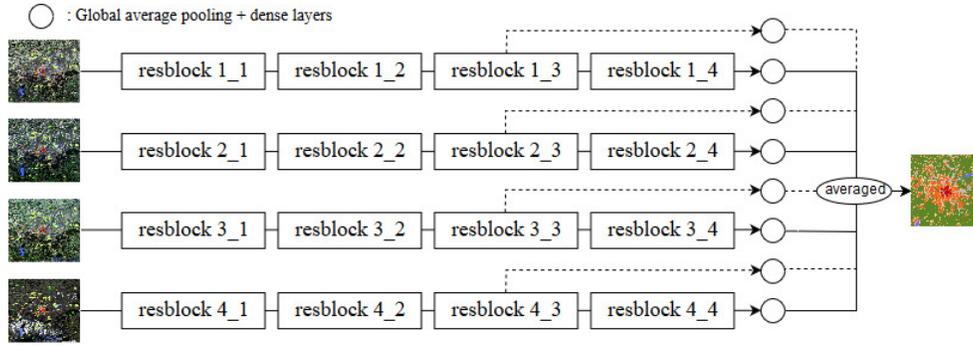


Fig. 1: End-to-end fusing network architecture for urban land cover classification, where T_1, T_2, T_3, T_4 are four seasons. The dropout is not shown. The network is referred as Res_ensemble_8 and Res_ensemble_4, when the dotted lines are considered and not considered, respectively.

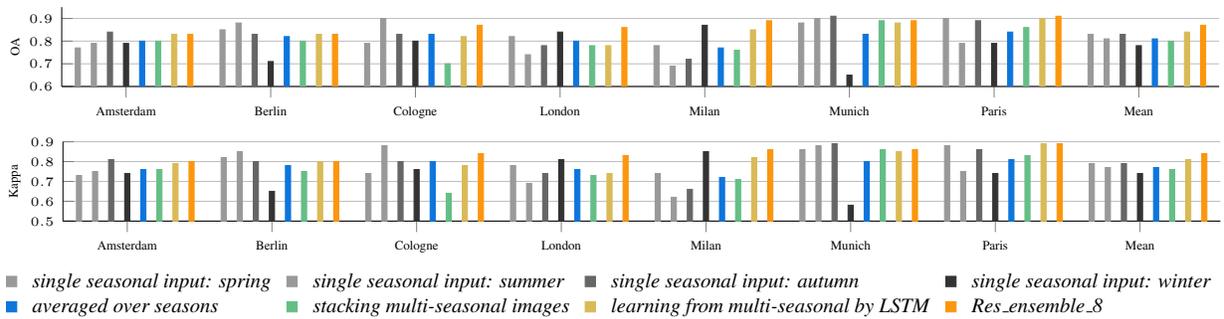


Fig. 2: OA and Kappa resulted from different approaches. For each city, training samples are from the other six cities and the last column (mean) is the averaged results over all 7 test cases, corresponding to the values in Tab. 1.

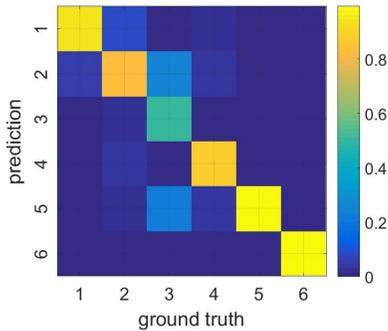


Fig. 3: Combined confusion matrices of the seven test cases.

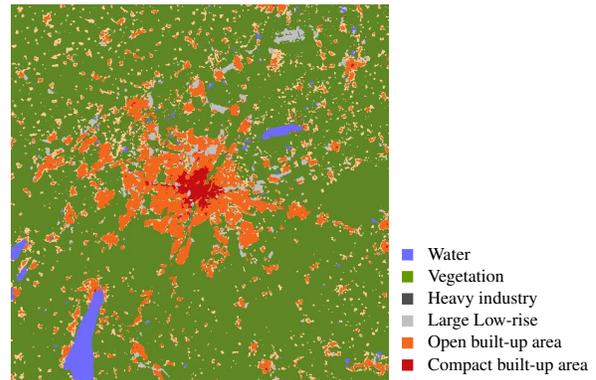


Fig. 4: Land cover map of the Munich city. The GSD is 100 meter. The shown area is of the size 600×600 pixel.

5, they also demonstrate some problems, which cannot be reflected by the accuracy measures. When enlarging the green rectangles in Fig. 5, it can be seen that some small buildings are unfortunately misclassified into Vegetation, while they are supposed to be classified into Sparsely built (cf. Fig. 6). This can also be seen from the confusion matrix shown in Fig. 3. This is due to the size of the building, which also explains why it is also missing in the GHS built-up grid dataset. Additionally, it can be a result of the employed patch-based clas-

sification. Patch-based classification is also one of the reasons that the produced land cover map is not as detailed as the GHS built-up grid, as can be seen in Fig. 5. Further improvement can be done regarding the input patch size of the network. Also, it can be beneficial to include some auxiliary datasets such as built-up grid during the training of the net-

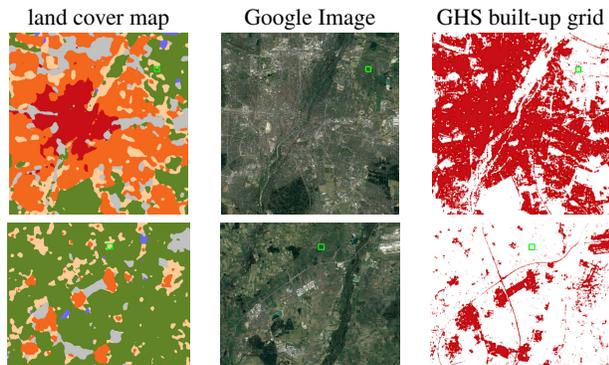


Fig. 5: Produced land cover maps, the optical image taken from Google Earth, and the GHS built-up grid, in the city center (up) and suburban (down) of Munich, Germany. The GSD of the land cover map is 10 meter. The red area in the GHS dataset indicates built-up area till the year 2014, with a resolution of 38 meter. The green rectangle areas will be further enlarged in the discussion part. The satellite image data: Google, Image Landsat / Copernicus.

work, which might guide the network to pay attention on the built-up areas.

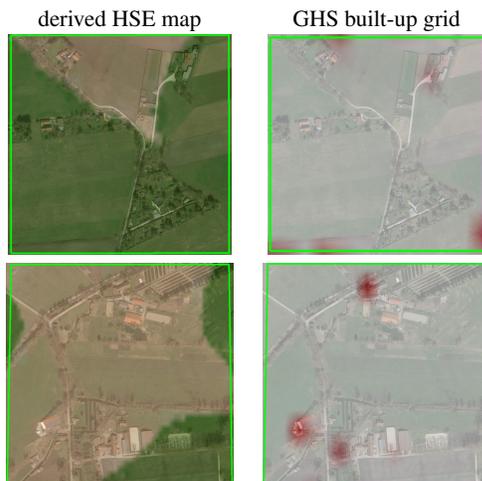


Fig. 6: Zoomed-in view of the green rectangle areas (in Tab. 5) overlaid to the Google earth images. The satellite image data: Google, Image Landsat / Copernicus.

5. CONCLUSION AND OUTLOOK

LCZ-derived urban land cover maps can be used as a first step for hierarchical LCZ classification or urban human settlement extent mapping. Based on our previous studies, this paper proposes a framework to fuse multi-seasonal Sentinel-2 images in a decision level. A set of cross-validations under the

same experimental setup, over a seven-city study area in central Europe demonstrates its better performance over several previously published approaches. Together with our previous work, we can conclude that decision-level fusion is better than feature-level fusion for similar tasks at similar scale, when multi-seasonal Sentinel-2 images are exploited. This work adds insights into similar types of urban mapping when using this globally openly available Sentinel-2 images.

6. ACKNOWLEDGEMENTS

This work is jointly supported by the China Scholarship Council (CSC), the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation program (Grant Agreement No. ERC-2016-StG-714087, Acronym: So2Sat), the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018, www.sipeo.bgu.tum.de) and the Bavarian Academy of Sciences and Humanities in the framework of Junges Kolleg.

7. REFERENCES

- [1] I. D. Stewart, "Local climate zones: Origins, development, and application to urban heat island studies," in *Proc. Annual Meeting of the American Association of Geographers, Seattle, Washington, USA*, 12–16 April, 2011.
- [2] C Qiu, M Schmitt, L Mou, P Ghamisi, and X X Zhu, "Feature Importance Analysis for Local Climate Zone Classification Using a Residual Convolutional Neural Network with Multi-Source Datasets," *Remote Sensing*, vol. 10, no. 10, pp. 1572, 2018.
- [3] M. Demuzere, B. Bechtel, and G. Mills, "Global transferability of local climate zone models," *Urban Climate*, vol. 27, pp. 46–63, 2019.
- [4] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. L. Saux, G. Moser, and D. Tuia, "Open Data for Global Multimodal Land Use Classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1363–1377, 2018.
- [5] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "LCZ-based urban land cover classification from multi-seasonal sentinel-2 images with a recurrent residual network," 2018, to appear.
- [6] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, et al., "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.
- [7] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 4, pp. 6–23, 2016.
- [8] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017.
- [9] X. X. Zhu and et al, "So2Sat LCZ42: A Benchmark Dataset for Global Local Climate Zones Classification," to appear, 2019.
- [10] M. Pesaresi, D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris, "Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014," *Publications Office of the European Union*, pp. 1–62, 2016.
- [11] M. Schmitt, L. Hughes, C. Qiu, and X. X. Zhu, "Aggregating Cloud-Free Sentinel-2 Images with Google Earth Engine," to appear, 2019.