Deep Convolutional Neural Networks for Semantic Segmentation of Multispectral Sentinel-2 Satellite Imagery: An Open Data Approach to Large-Scale Land Use and Land Cover Classification

LUKAS LIEBEL[1]

Abstract: Large-scale land use and land cover (LULC) maps with up to global coverage still incorporate large amounts of manual work and are thus only available after production times of several years. In the domain of computer-vision, deep convolutional neural networks (CNNs) dominate the state-of-the-art in image recognition, a task closely related to LULC classification. By adopting a state-of-the-art CNN architecture for the classification of multispectral Sentinel-2 image patches, the close relation of both tasks is exploited. Since no suitable dataset for training is available as of now, a quasi-automatic method for deriving labels from existing sources of volunteered geographic information, in particular the OpenStreetMap, is presented. The results indicate that CNNs for image recognition are well-suited for LULC classification with only minor adaptions and are furthermore robust to label noise, originating from the error-prone data acquisition method.

1 Introduction

One of the core fields in earth observation is gathering information about the surface of the earth. The acquired data allows for a better understanding of our living environment and thus serves as a basis for decision making in numerous domains. Spaceborne remote sensing enables applications on a global scale. A typical product created for such applications are *land use and land cover* (LULC) maps, which classify the covered area into regions of similar characteristics regarding their surface coverage and employed usage.

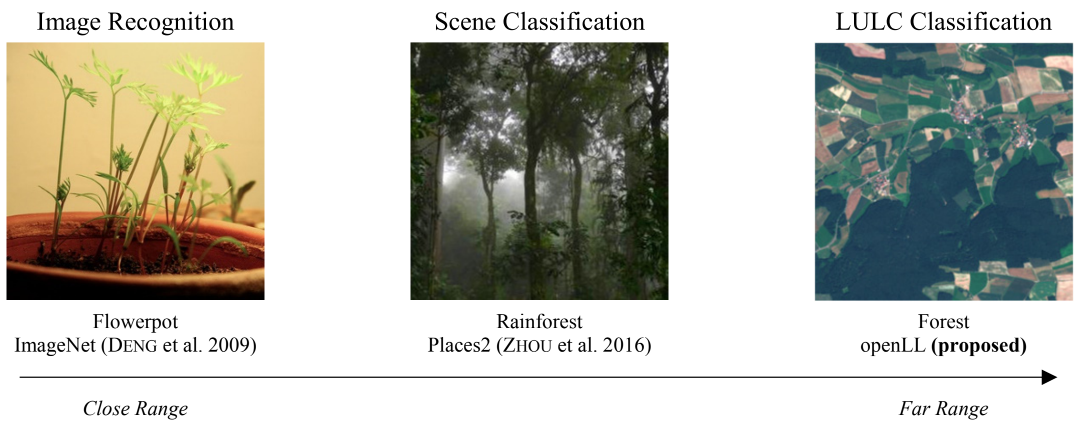

Fig. 1: Relation of image classification tasks from the computer vision domain and LULC classification

[1] Technical University of Munich, Chair of Remote Sensing Technology, Arcisstr. 21, D-80333 München, E-Mail: lukas.liebel@tum.de

Automatic methods for LULC classification exist. However, they often struggle with the difficulties of large-scale LULC classification, in particular the low intra-class and high inter-class similarity.
Pixel-wise spectral features, as used by most traditional methods, are often not expressive enough to cover the variance of the target classes. Recent approaches, thus, consider spatial context by using image regions or patches for classification instead of single pixels. Utilizing image patches for classification hints a close relation to the tasks image recognition and scene classification from the computer vision domain, as illustrated in Fig. 1.

2 Approach

Techniques based on deep learning were able to achieve unmatched results for image recognition in the last years. As LULC classification is closely related to this task, existing methods can be adopted. This can potentially help to overcome shortcomings of the currently employed traditional classification methods. *Convolutional neural networks* (CNN), which represent the state-of-the-art in image recognition, are well-known for being able to generalize well to unseen samples, even in very fine-grained classification. However, CNN architectures for image recognition cannot directly be applied to LULC classification, since remote sensing imagery differs from common photographs in several decisive properties, such as spatial, spectral, and radiometric resolution. Few remote sensing datasets exist, which have been used for adapting image recognition CNNs for LULC classification (PENATTI et al. 2015; CASTELUCCIO et al. 2015). They are, however, not suitable for training a deep CNN for large-scale LULC from scratch, since this task comprises solving optimization problems with millions of parameters.
In order to still be able to evaluate the applicability of this promising technique to the given task, a comprehensive dataset for training was created. Since manual annotations are prohibitively expensive, a method for deriving image patches and ground-truth labels from open data is presented. Multispectral image patches and cloud masks were acquired through the *Copernicus* earth observation program, in particular, the *Sentinel-2* mission. Corresponding ground truth annotations for the creation of the proposed dataset *Open LULC Labels* (openLL) were derived from the *OpenStreetMap* (OSM) database. As semantic information is provided by existing crowdsourced data, the proposed method is considered a quasi-automatic annotation approach. This concept allows for scaling the approach to an arbitrarily large area and number of samples with constant and almost negligible manual effort.
The proposed dataset enables for training a CNN for large-scale LULC classification. With both state-of-the-art performance and available source code for the *TensorFlow* framework, the *Inception-v3* CNN architecture (SZEGEDY et al. 2016) is well-suited and was thus chosen for the experiments. The architecture was adapted to the requirements of multispectral Sentinel-2 image patches and optimized on the training set.

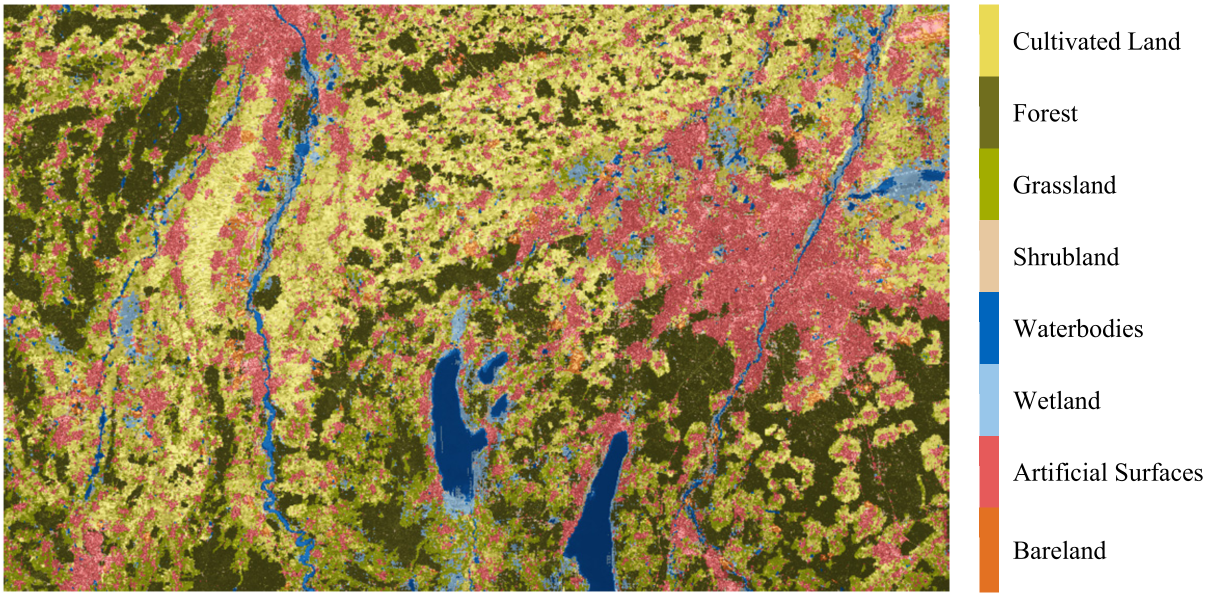

Fig. 2: Land use and land cover map produced by the proposed CNN-based classification approach

3 Results and Conclusion

The classification performance of the trained CNN was assessed using the manually annotated test set of openLL. As a reference, the results were compared to existing LULC maps, namely *GlobeLand30* (GLC30; CHEN et al. 2015) and *CORINE Land Cover* (CLC), which represent the state-of-the-art but involve a large amount of manual work. They achieved a recall of 36% and 44%, and a high overall accuracy of 76.82% and 79.88%, respectively, which was expected. The CNN classifier was applied by using a sliding window approach. With an average recall of 45% and an overall accuracy of 50.72%, the presented approach achieved solid results. Fig. 2 shows a qualitative classification result, enhanced by the suggested multi-look classification procedure. Some of the remaining misclassifications can be traced back to issues in the translation of OSM tags to LULC labels. Further experiments were conducted on the robustness of the classifier to label noise, introduced by the error-prone method employed for creating training data. In addition to that, the propagation of errors from faulty training data to classification results was explored.

For a selected class (clouds opaque), dedicated test areas were manually annotated using a binary classification scheme. A direct comparison to the Sentinel-2 Level-1C cloud masks, used as training data, resulted in a recall of only 5.29% and an overall accuracy of 74.15% for the original cloud masks, and a recall of 21.26% and an overall accuracy of 78.29% for the CNN classification result. The CNN thus proved to be able to cope with conflicting training data.

These results support the hypothesis that an image recognition CNN can successfully be re-trained for the LULC classification task using quasi-automatically annotated training data from open data sources. Since the current implementation is a mere feasibility analysis and is not optimized for achieving competitive classification performance, suggestions for enhancing the proposed approach are given.

4 References

CASTELLUCCIO, M., POGGI, G., SANSONE, C. & VERDOLIVA, L., 2015: Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. Online: arXiv:1508.00092 [cs.CV] (last accessed: 18-01-2017).

CHEN, J., CHEN, J., LIAO, A., CAO, X., CHEN, L., CHEN, X., HE, C., HAN, G., PENG, S., LU, M., ZHANG, W., TONG, X. & MILLS, J., 2015: Global Land Cover Mapping at 30 m Resolution: A POK-based Operational Approach. ISPRS Journal of Photogrammetry and Remote Sensing, **103**, 7-27.

DENG, J., DONG, W., SOCHER, R., LI, L. J., LI, K. & FEI-FEI, L., 2009: ImageNet: A Large-scale Hierarchical Image Database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 248-255.

PENATTI, O. A. B., NOGUEIRA, K. & DOS SANTOS, J. A., 2015: Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 44-51.

SZEGEDY, C., VANHOUCKE, V., IOFFE, S., SHLENS, J. & WOJNA, Z., 2016: Rethinking the Inception Architecture for Computer Vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818-2826.

ZHOU, B., KHOSLA, A., LAPEDRIZA, A., TORRALBA, A. & OLIVA, A., 2016: Places: An image database for deep scene understanding. Online: arXiv preprint arXiv:1610.02055 [cs.CV] (last accessed: 18-01-2017).