Returning the favor - Leveraging quality insights of OpenStreetMap-based land-use/land-cover multi-label modeling to the community

Moritz Schott[1,*], Adina Zell[2], Sven Lautenbach[1,3], Begüm Demir[2] and Alexander Zipf[1,3]

[1] Institute of Geography/GIScience, Heidelberg University, Heidelberg, Germany; moritz.schott@uni-heidelberg.de, sven.lautenbach@uni-heidelberg.de, zipf@uni-heidelberg.de
[2] Faculty of Electrical Engineering and Computer Science/Remote Sensing Image Analysis Group, Technische Universität Berlin, Berlin, Germany; adina.zell@campus.tu-berlin.de, demir@tu-berlin.de
[3] HeiGIT, Heidelberg University, Heidelberg, Germany

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2022 Conference after peer-review.

Land-use and land-cover (LULC) information in OSM is a challenging topic. On the one hand, this information provides the background for all other data rendered on the central map and is used by applications like https://osmlanduse.org. It has a high potential to be used as a valuable data source to tackle major current challenges like the climate crisis. On the other hand, this information has a difficult position within the OSM ecosystem. LULC information can be quite cumbersome or even difficult to map e.g. due to natural ambiguity. As most other OSM tagging schemes, the current LULC tagging scheme is the result of a bottom-up growth which resulted in a collection of sometimes ambiguous, unstable or overlapping tag definitions that are not fully compatible with any official LULC legend definition [1]. Furthermore, the data is highly shaped by local or national preferences and imports. This diversity of the LULC data in OSM is a fundamental principle of OSM that enabled the success of the project. Yet, this can create considerable usage barriers or at least caveats for data users unfamiliar with the projects' ecosystem. The remote sensing community for instance has started to use OSM LULC information as labels in their classification models. Frequently, OSM LULC data has thereby been taken at face value without critical reflection. And, while the quality and fitness for purpose of OSM data has been proven in many cases (e.g., [2,3]), these analyses have also unveiled quality variations e.g. between rural and urban regions. The quality of OSM therefore can be assumed to be generally high, but remains unknown for a specific use-case.

The proposed work first assesses the impact of these challenges on a use-case of multi-label remote sensing (RS) image classification and then provides a machine learning (ML) based workflow to overcome and finally mitigate them. RS images contain multiple

Schott, M., Zell, A., Lautenbach, S., Demir, B., & Zipf, A. (2022). Returning the favor - Leveraging quality insights of OpenStreetMap-based land-use/land-cover multi-label modeling to the community.
In: Minghini, M., Liu, P., Li, H., Grinberger, A.Y., & Juhász, L. (Eds.). Proceedings of the Academic Track at State of the Map 2022, Florence, Italy, 19-21 August 2022. Available at https://zenodo.org/communities/sotm-22
DOI: 10.5281/zenodo.7004593

LULC classes and thus can be simultaneously associated with different class labels (i.e. multi-labels). The development of multi-label RS image classification methods, which aim at automatically assigning class labels to each RS image in an archive, is a growing research interest. In the presented study multi-labels are extracted from OSM and used to train the ML model.

The fitness for purpose of OSM for multi-label RS image classification was tested on a Sentinel 2 scene with a resolution of 10m and four bands in south west Germany recorded in June 2021. The area was chosen for its estimated high completeness of OSM land use information and for the low amount of imported data. OSM data was grouped by its tags into the four LULC classes 'forest', 'agricultural area', 'build-up area' and 'water body'. 18 tags that could unequivocally be mapped to these classes were used and small elements, below the image resolution or the classes minimal mapping unit, were filtered. The chosen scene was then tiled into a 1.22 x 1.22 km grid of 8100 image patches. Zero to four labels were assigned to each patch based on the OSM LULC elements therein. Evaluation was performed manually on 910 random patches, of which 80% had a correct OSM-based multi-label, thereby proving the assumed high completeness and quality in the region.

The proposed workflow provides a method to enhance this OSM-based RS image multi-label classification and extend it to areas of lower OSM quality and completeness using ML, specifically deep learning (DL). The main obstacle for ML and especially DL is the required amount of labeled training data. Volunteered geographic information (VGI) like OSM offers a potential solution to this challenge by providing an overabundance of LULC information that is suitable for this purpose - if data quality is sufficiently high. The workflow uses the multi-label information extracted from OSM for training and then detects discrepancies between its predictions and OSM. Using this information and pinpointing the exact location of error within the patches provides valuable OSM data quality information.

Apart from facilitating a fast quality estimation for large areas, the workflow can also make its findings automatically available to the OSM community in a feedback loop using the HOT Tasking Manager framework. Thereby, the valuable service by the OSM community of providing large amounts of free and generally high quality training data is recognised in the form of quality feedback including mapping hints to the OSM community. The five workflow stages are: 1) RS data collection and preprocessing, 2) OSM data collection and preprocessing, 3) LULC multi-label modeling, 4) OSM data issue flagging and 5) the community feedback loop via Tasking Manager projects. While each step is an atomic use case and application, the combination of all four steps creates a tool that is useful for the RS and the OSM community likewise. The tool is openly available at https://gitlab.gistools.geog.uni-heidelberg.de/giscience/ideal-vgi/osm-multitag under the GNU Affero General Public License v3 including example datasets. Manual input was kept as low as possible while enabling the 'human in the loop' to take full control over all input and output.

The workflow extracts multi-label training data in stages 1) and 2) as described. Stage 3) then trains a DL model to: i) predict multi-labels; ii) identify incorrect (training) labels; and iii) localise the areas associated with these incorrect labels. Multi-labels can be incorrect if either one or more labels are missing (omission) or labels are wrongly assigned (commission). Omission errors occur when data is missing in OSM while commission errors occur when OSM data is falsely mapped in the tile. Multi-labels that are only partly incorrect, e.g. containing only a single wrong label, are often called ‘noisy labels’.

A DeepLabv3+ model [4] where the segmentation head is replaced with a multi-label classification head [5] is used as classification model. Yet, the output of the model does not provide spatial information regarding the location of the LULC classes present in the images. To localise the classes in a given image, we investigate the effectiveness of explainable neural networks. Explanation methods are capable of generating explanations in the form of pixel-level heatmaps that can be highly relevant for providing class localisation maps from a DL model trained using image level labels. In this work, we exploit self-enhancement maps [6] due to its proven success in providing accurate explanations. The class localisation is then combined with error detection techniques allowing the visualisation of potentially incorrect labels from the OSM data in stage 4). Depending on the availability of verified noise-free labels, one of two different approaches for noise detection is applied. If no clean data is available, the model class prediction is directly compared with the corresponding label from OSM. If noise-free data is available, an adapted noise detection method of CleanNet [7] for multi-labels is applied. CleanNet generates a single representative class prototype for each class and uses it to estimate the correctness of sample labels. Thereby, the model is first trained on clean data. Afterwards, for each class, a prototype is extracted based on the feature maps of the noise-free data. To detect if an image has wrong labels, the extracted features of the image with potentially noisy labels are compared with the corresponding class prototypes. The similarity between the two indicates whether the OSM label is potentially noisy.

For demonstration, the model was trained on the described Sentinel 2 scene in Germany. The models' performance was validated on the manually labelled 910 patches. The model achieved a mean average precision (MAP) of 0.98, whereas the direct use of OSM tags yields a MAP of 0.91. For the selected test region, the OSM data quality was very high and the above described filter procedure assured a balanced amount of omission and commission errors. To assess the effectiveness of the model under higher label-noise rates, simulating areas of lower OSM quality or completeness, synthetic noise was added to the labels using the approach established by [8]. This ensures that both omission and commission label noise are introduced to the multi-label training set. When synthetically decreasing the OSM training data quality to a MAP of 0.78, the model maintained a MAP above 0.97. Models trained on further deteriorated training data were also able to achieve surprisingly high prediction rates. Detailed investigation though showed that these models had moved towards a probabilistic approach of predicting classes more and more based on their occurrence in the training set rather than the actual image features.

The experiment shows that for DL approaches, label quantity is often of higher importance compared to label quality. Given the sheer overabundance of OSM training labels, OSM data quality can therefore be seen as a secondary problem, especially regarding the fact that it is generally high and can partly be assessed and assured through prior data processing. It can also be stated that OSM noise detection using DL is possible even for models trained on areas with relatively low OSM quality and completeness. Yet, the separation of stages 3) and 4) also allows the application of pretrained models in case OSM data is completely missing or suspected to be below a usable quality threshold. While this application is under active research, it is suspected that pretrained models from regions with higher OSM data quality can yield good multi-label prediction and noise detection results in regions with very low OSM quality and completeness. Provided the existence of a comparable region in terms of class distribution, definition and appearance, this would make the workflow widely applicable to most global regions or historic timestamps, independent of local OSM data.

The final stage 5) uses the above described localised potential OSM data errors to create HOT Tasking Manager projects via the API. These projects provide additional correction hints. Yet, no automatic editing takes place. The community is kept in full control of all mapping actions as a 'human in the loop'.

The high fitness for purpose of OSM has been proven for the use-case of multi-label RS image classification. The proposed tool provides an automated OSM multi-label extraction, modeling and verification procedure including a return of results to the OSM community. A major challenge of the approach is the tiled view on the data. If OSM assigns correct multi-labels to a patch, more fine grained data issues will not be detected. Yet, this approach allows large scale data assessments, before the topic of more detailed data improvement is tackled. It also allows to run repeated OSM LULC quality and completeness estimations for large areas over time including far reaching retrospectives. Another major benefit is the usage of local OSM data for modeling, thus making regionalised models the standard procedure. This is required for OSM LULC information as regional data structures and communities exist, which need to be preserved. The model can lead to regional homogenisation and data cohesion within these regional communities.

Acknowledgements: This work was funded by DFG, German Research Foundation ("IdealVGI", 424966858).

References

[1] Fonte, C., Minghini, M., Antoniou, V., See, L., Patriarca, J., Brovelli, M.A. and Milcinski, G.(2016). An automated methodology for converting OSM data into a land use/cover map. In: Bandrova, T,. & Konecny, M. (Eds) *Proceedings of the 6th International Conference on Cartography & GIS*, Albena, Bulgaria, 13-17 June 2016, 482–473.

[2] Jokar Arsanjani, J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality Assessment of the contributed land use information from openstreetmap versus authoritative datasets. In: Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (Eds.) *OpenStreetMap in GIScience*, Springer, Cham, 37–58.

[3] Fonte, C.C., Bastin, L., See, L., Foody, G., & Lupia, F. (2015). Usability of VGI for validation of land cover maps. *International Journal of Geographical Information Science, 29*(7), 1269–1291.

[4] Chen, L., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., & Weiss, Y. (Eds.) *Computer Vision – ECCV 2018*, Springer, Cham, 833–851.

[5] Ahlswede, S., Madam, N.T., Schulz, C., Kleinschmit, B., & Demr, B. (2022). Weakly supervised semantic segmentation of remote sensing images for tree species classification based on explanation methods. *arXiv preprint*, arXiv:2201.07495.

[6] Zhang, X., Wei, Y., Yang, Y., & Wu, F. (2020). Rethinking Localization Map: Towards Accurate Object Perception with Self-Enhancement Maps. *arXiv preprint*, arXiv:2006.05220v2.

[7] Lee, K., He, X., Zhang, L., & Yang, L. (2018). CleanNet: Transfer Learning for Scalable Image Classifier Training with Label Noise. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5447–5456.

[8] Burgert, T., Ravanbakhsh, M., & Demir, B. (2022). On the effects of different types of label noise in multi-label remote sensing image classification. *arXiv preprint*, arXiv:2207.13975.