Learning Building Floor Numbers from Crowdsourced Street-view Images

Yifan Tian*,Yao Sun*, Xiao Xiang Zhu

Data Science in Earth Observation, Technical University of Munich, (yifan.tian, yao.sun, xiaoxiang.zhu)@tum.de

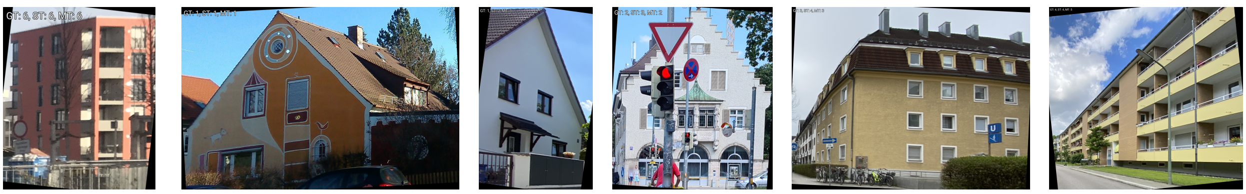
Keywords: Building Floor Number, Deep Learning, Street-view Images (SVI), Mapillary, Crowdsource, Large Scale.

Abstract:

Building information extraction has been a hot topic for three decades. However, most of the research efforts primarily focus on building geometries, e.g., Li et al. (2020), Sun et al. (2019, 2022), Chen et al. (2021), but not on attributes. One important building attribute is the floor number, which is crucial for urban planning, aiding in estimating household numbers, energy consumption, renovation costs, property management, and emergency responses. However, this data is often unavailable and cannot be directly inferred from 3D models or building heights, as pointed out in Roy et al. (2023). While remote sensing images are helpful for height estimation, they aren't suitable for floor number estimation due to their nadir view. Street-view imagery (SVI) is a good data source for this task as it features building facades. A few studies use SVI from commercial platforms such as Google Street View, e.g., Iannelli and Dell'Acqua (2017), Wu et al. (2021), Kramm et al. (2023). However, due to the limitations on the data license, challenges exist in applying these approaches on a larger scale. Besides those offered by commercial services, SVIs can also be crowdsourced by citizens and managed by platforms such as Mapillary or Flicker, and contribute to building information extraction, e.g., Hoffmann et al. (2023), Sun et al. (2023). However, compared to commercial counterparts, processing crowdsourced images poses challenges, including variations in image quality and metadata accuracy.

This work presents a framework for large-scale building floor number estimation with two main contributions: 1) a dataset generation pipeline that creates an SVI building dataset, and 2) a multi-task learning deep neural network that incorporates roof information to enhance floor number estimation accuracy.

The dataset generation pipeline creates "ImageCrops" from Mapillary images for each building through three steps. First, buildings are detected and cropped from SVIs using the Grounding DINO object detection model, as developed in Liu et al. (2023). Second, a binary search algorithm based on the SVI's field of view (FoV) matches cropped images to building footprints. Third, semantic segmentation and statistical analysis between segments filter out ImageCrops that do not fully depict the building or include occluding non-building objects. Finally, manual checks ensure dataset quality. Our curated dataset in Munich comprises 6,473 images for 4,129 buildings, with building footprints and floor number ground truth sourced from official Munich building models[1].



(a) 6, 6, 6, 6 (b) 1, 1, 1, 2 (c) 1, 2, 1, 2 (d) 2, 3, 2, 3 (e) 3, 4, 3, 4 (f) 4, 4, 5, 3

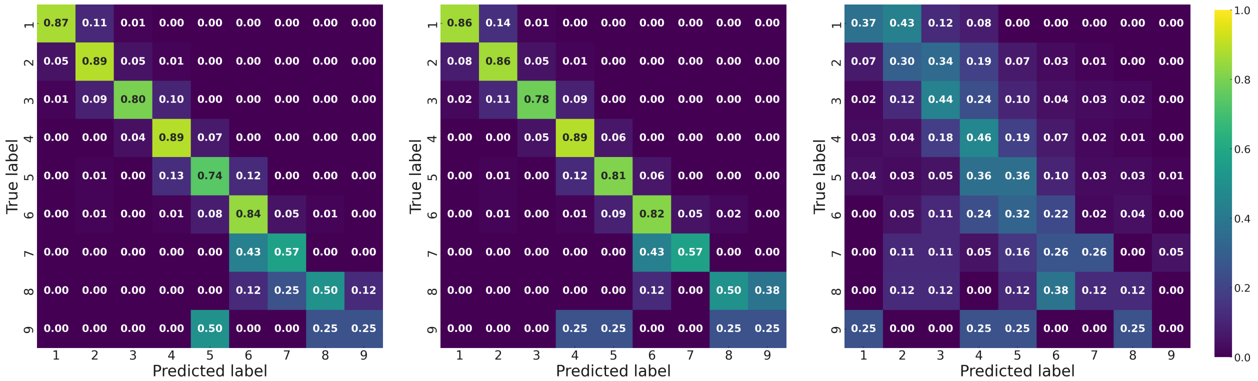
Figure 1. Examples of classification results. The floor numbers of GT, STL, MTL, and Clustering results are highlighted in green, red, blue, and brown, repectively.

Building floor number estimation is formulated as a classification task, where we developed a multi-task learning (MTL) to predict both the floor number and roof type of buildings. Comparative experiments are single-task learning (STL) using to classify building floors only, and a clustering-based approach that vertically clusters detected windows to regress the floor number. The MTL model utilizes ResNet-50 with Adaptive Average Pooling and spatial pyramid pooling. For dynamic expert selection, it employs DSelect-k, and balances losses using Random Loss Weighting. We evaluate the results using overall accuracy and normalized confusion matrices (c.f., Figure 2), and some results are shown in Figure 1. The MTL and STL models perform well overall, with MTL showing superior performance in handling buildings with

* Equal contribution

[1] https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=lod2

https://doi.org/10.5194/ica-abs-7-171-2024



(a) MTL (b) STL (c) Clustering

Figure 2. Normalized confusion matrices for MTL model, STL model, and clustering approach.

diverse roofs. The MTL model achieves an overall accuracy of 84.22%, a notable improvement of 0.62% over the STL model's accuracy of 83.60%. In contrast, the clustering-based approach achieves an overall accuracy of only 36.81%. The MTL model excels particularly in accurately predicting buildings with non-flat roofs, where the STL model has limitations, underscoring the effectiveness of the MTL approach.

Despite the satisfactory performance, the quantity and quality of the building ImageCrops is the bottleneck for large-scale building floor number estimation. The current dataset lacks diversity, covering only floor numbers 1 to 9. Future improvements will focus on expanding dataset variability.

Acknowledgements

The work is jointly supported by by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 499168241 and by the Technical University of Munich (TUM) Georg Nemetschek Institute under the project Artificial Intelligence for the automated creation of multi-scale digital twins of the built world (Acronym: AI4TWINNING).

References

Chen, S., Mou, L., Li, Q., Sun, Y. and Zhu, X. X., 2021. Mask-height R-CNN: An end-to-end network for 3D building reconstruction from monocular remote sensing imagery. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*.

Hoffmann, E. J., Abdulahhad, K. and Zhu, X. X., 2023. Using social media images for building function classification. *Cities* 133, pp. 104–107.

Iannelli, G. C. and Dell'Acqua, F., 2017. Extensive exposure mapping in urban areas through deep analysis of street-level pictures for floor count determination. *Urban Science* 1(2), pp. 16.

Kramm, A., Friske, J. and Peukert, E., 2023. Detecting floors in residential buildings. In: *German Conference on Artificial Intelligence (Künstliche Intelligenz)*, Springer, pp. 130–143.

Li, Q., Mou, L., Hua, Y., Sun, Y., Jin, P., Shi, Y. and Zhu, X. X., 2020. Instance segmentation of buildings using keypoints. In: *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J. et al., 2023. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*.

Roy, E., Pronk, M., Agugiaro, G. and Ledoux, H., 2023. Inferring the number of floors for residential buildings. *International Journal of Geographical Information Science* 37(4), pp. 938–962.

Sun, Y., Hua, Y., Mou, L. and Zhu, X. X., 2019. Large-scale building height estimation from single VHR SAR image using fully convolutional network and GIS building footprints. In: *Joint Urban Remote Sensing Event (JURSE)*.

Sun, Y., Kruspe, A., Meng, L., Tian, Y., Hoffmann, E. J., Auer, S. and Zhu, X. X., 2023. Towards large-scale building attribute mapping using crowdsourced images: scene text recognition on flickr and problems to be solved. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLVIII-1/W2-2023, pp. 225–232.

Sun, Y., Mou, L., Wang, Y., Montazeri, S. and Zhu, X. X., 2022. Large-scale building height retrieval from single sar imagery based on bounding box regression networks. *ISPRS Journal of Photogrammetry and Remote Sensing* 184, pp. 79–95.

Wu, M., Zeng, W. and Fu, C.-W., 2021. Floorlevel-net: recognizing floor-level lines with height-attention-guided multi-task learning. *IEEE Transactions on Image Processing* 30, pp. 6686–6699.

https://doi.org/10.5194/ica-abs-7-171-2024 |