Bavaria Buildings - A Novel Dataset for Building Footprint Extraction, Instance Segmentation, and Data Quality Estimation (Data and Resources Paper)

Martin Werner
Professorship of Big Geospatial Data Management, Technical University of Munich, Germany
martin.werner@tum.de

Hao Li
Professorship of Big Geospatial Data Management, Technical University of Munich, Germany
hao_bgd.li@tum.de

Johann Maximilian Zollner
Professorship of Big Geospatial Data Management, Technical University of Munich, Germany
maximilian.zollner@tum.de

Balthasar Teuscher
Professorship of Big Geospatial Data Management, Technical University of Munich, Germany
balthasar.teuscher@tum.de

Fabian Deuser
Institute for Distributed Intelligent Systems, University of the Bundeswehr Munich, Germany
fabian.deuser@unibw.de

ABSTRACT

Bavaria Buildings is a large, analysis-ready dataset providing openly available co-registered 40cm aerial imagery of Upper Bavaria paired with building footprint information. The Bavaria Buildings dataset (BBD) contains 18205 orthophotos of 2500×2500 pixels, where each pixel covers 40cm × 40cm in space (Digitales Orthophoto 40cm - DOP40). The dataset has been pre-processed and co-registered and also provides a set of 5.5 million image tiles of 250×250 pixels ready for deep learning and image analysis tasks. For each image tile, we provide two segmentation masks; one based on the official building footprints (Hausumringe) data as published by the Free State of Bavaria and one based on a historic OpenStreetMap (OSM) extract dating to 2021. The dataset is ready for essential analysis tasks, such as detection, segmentation, instance extraction, footprint geometry extraction, multimodal localization, and multimodal data quality assessment of buildings in Bavaria. We plan to update the dataset with each major re-publication of the upstream data sources to foster change detection research in the future. The BBD is available at https://doi.org/10.14459/2023mp1709451.

CCS CONCEPTS

• **Information systems → Geographic information systems**; • **Computing methdologies → Artificial intelligence**.

KEYWORDS

Dataset, Orthophoto, Very High Resolution, Building Detection, OpenStreetMap, Data Quality, Geospatial Artificial Intelligence

SIGSPATIAL '23, November 13–16, 2023, Hamburg, Germany

ACM ISBN 979-8-4007-0168-9/23/11…$15.00
https://doi.org/10.1145/3589132.3625658

ACM Reference Format:
Martin Werner, Hao Li, Johann Maximilian Zollner, Balthasar Teuscher, and Fabian Deuser. 2023. Bavaria Buildings - A Novel Dataset for Building Footprint Extraction, Instance Segmentation, and Data Quality Estimation (Data and Resources Paper). In *The 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23), November 13–16, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3589132.3625658

1 INTRODUCTION

Observing the Earth's surface from airborne imaging sensors is a long-standing tradition in many countries. Authorities use such data to update the official information products and to compare reality with the governmental databases available. Furthermore, in situations such as natural disasters like Earthquakes or floods, such images are comparably easy to acquire and provide a rich data source. Unfortunately, such data is not commonly shared with the public. However, the Free State of Bavaria, Germany, recently made a 40cm resolution high-quality orthophoto dataset freely available under an open data license. Besides Very High Resolution (VHR) images, Bavaria has published additional datasets, including official maps, a DEM, and others. Most interestingly, all buildings, as modeled in the governmental survey of Bavaria, are made available as simple and easy-to-use shapefiles. While these two datasets alone might not be entirely in sync (e.g., the images might show buildings that are not in the official data and vice versa), the official data can contain buildings that are not yet built or have lately been destroyed. A similar situation appears when considering a third building dataset extracted from the freely available OpenStreetMap (OSM) data. OSM tends to be fast when there is a good motivation for excessive mapping (e.g., in disaster response [7]). However, in some areas, OSM is based on out-of-date imagery, so we expect certain differences between the VHR image dataset, the OSM, and the governmental building dataset. Moreover, VHR images are mostly not available in similar existing datasets (e.g., Microsoft building footprints[1]), thus limiting the downstream data applications to a

[1]https://github.com/microsoft/GlobalMLBuildingFootprints

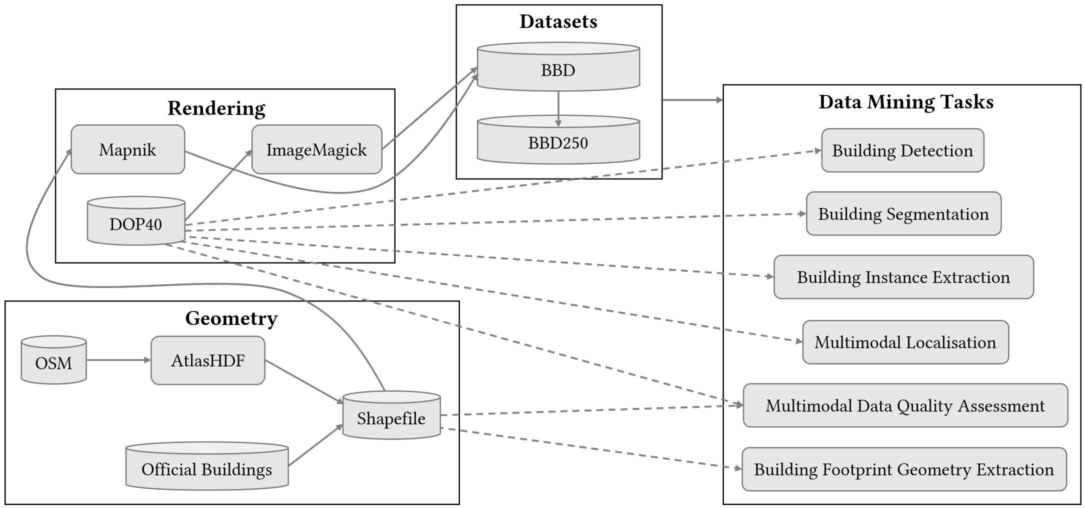

Figure 1: Overview of the BBD datasets.

certain extent.

In this paper, we collect this triple of data sources as a curated dataset, covering the whole area of the Upper Bavaria, Germany, in an analysis-ready presentation suitable for multiple communities, including computer vision, computational geometry, geographic information science, and machine learning.

2 INPUT DATASET DETAILS

2.1 DOP40 Orthophoto of Bavaria

The Bavarian mapping agency provides an orthophoto dataset in UTM Zone 32M projection covering the whole Free State of Bavaria in 40cm resolution. This dataset provides a high-quality VHR dataset taken from airborne camera observations. Hence, the dataset captures RGB color information about the Earth's surface with almost no atmospheric distortions but some good weather artifacts such as comparably dark shadows.

The spatial resolution of this dataset is good enough to delineate fine-grained structures such as streets and buildings in quite some detail. At the same time, it only has limited privacy issues, as many mobile objects of humans (e.g., cars), are challenging to identify.

The DOP40 dataset is originally available from the open data portal[2] in Bavaria using a metalink downloader process as a set of GeoTIFF files, each holding 2500×2500 RGB pixels in a common 24bit color space all files being in the aforementioned projection. In this manner, the DOP40 dataset results in 18205 tiles of 2500×2500 pixels (each 40cm $\times$ 40cm) in total.

2.2 Official Building Footprint

The official building footprints, so-called "Hausumringe", are shared as a shapefile in which every building is modeled as a polygon. Those shapefiles are available from the open data portal and can be individually downloaded for each county of Bavaria ("Regierungsbezirk"). These building footprints show the inner of georeferenced boundary polygons (vector data) of the building floor plans of the real estate cadastral plan from the state government, which can be regarded as an authoritative reference dataset.

2.3 The OpenStreetMap Buildings of Bavaria

OpenStreetMap (OSM) is a collaborative mapping project to create a digital map of the world [4]. OSM can offer higher precision and more semantic information on attributes than most RS products, although the volunteered geographical information (VGI) nature of OSM leads to inevitable concerns regarding data quality, position accuracy, spatial consistency, and data completeness [3, 5].

To extract the OSM building dataset, we download historic OSM data dumps dating back to 2021, then filter and extract all polygons with the key-value pair "$building = *$" using AtlasHDF [10]. As a result, we convert all buildings in Upper Bavaria into a Shapefile.

3 DATASET TRANSFORMATION AND GENERATION

Figure 1 shows an overview of preparing the BBD dataset from the aforementioned open data sources, namely DOP40, official building footprints, and OSM. In a geometry processing component, the OpenStreetMap database is filtered using AtlasHDF [10] and converted to a Shapefile of building polygons. Similarly, the various building polygons from the official government data platform are merged into a single polygon Shapefile. In the rendering stage, we convert geographic information into bare image information suitable for most deep learning frameworks; we use the Mapnik cartographic renderer[3] to render an image co-registered with each DOP40 geographic image (GeoTIFF). In addition, we convert each of these GeoTIFFs into a bare PNG file to generate simple image file pairs from all three datasets: a PNG image with identical image information as the given DOP40 image and two PNG images, one a co-registered rendering of OSM buildings and the other one a co-registered rendering of official building footprints.

[2] https://geodaten.bayern.de/opengeodata/

[3] https://mapnik.org/

In order to facilitate downstream deep learning tasks, the BBD dataset has been decomposed into even smaller tiles to reach an image size common to computer vision models. We use 250×250 pixels representing an area of $100m \times 100m$ in space. The tiles have a reasonable size for building detection and footprint segmentation and are, at the same time, small enough to accelerate deep learning model training with GPUs with larger batch sizes. Finally, this leads to a number of 5.5 million tiles of 250×250 pixels (around 210 GB) in the Upper Bavaria area, denoted as the BBD250 dataset.

4 CHARACTERIZATION OF THE BBD DATASET

In this paper, we provide a novel dataset, namely the BBD dataset, for researchers interested in geospatial data analysis, both with classical methods and modern methods based on Geospatial Artificial Intelligence (GeoAI). As a large-scale dataset, the BBD dataset represents a valuable resource for new research since the dataset goes beyond what previous datasets could offer for the community in the following aspects.

4.1 Spatial Autocorrelation

The dataset shows extreme spatial autocorrelation because buildings typically group in villages and cities. When picking a random image patch from the dataset, the probability of finding a patch that actually contains buildings is very low - most $100m \times 100m$ areas in Bavaria do not contain any buildings at all.

In the BBD250, for example, about 1.5 million out of the 5.5 million images (27%) do not contain a single building pixel. For the other images (containing at least one building pixel), we show the density of building pixels in the image in Figure 2. Herein, this spatial autocorrelation poses a new challenge for effective training and testing of AI-based building segmentation models in a real-world data imbalance scenario.

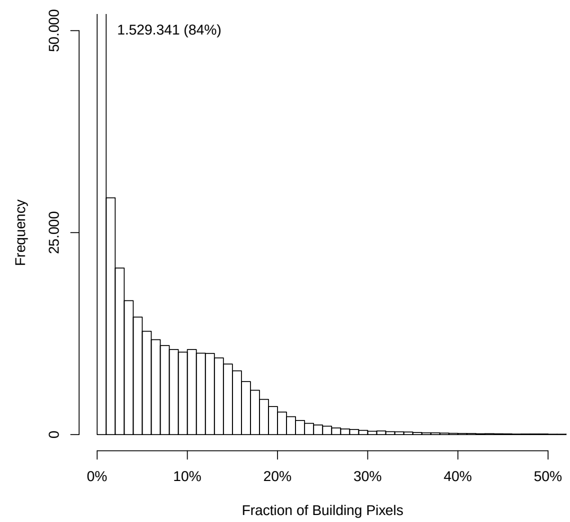

Figure 2: Histogram of building pixels per patch (BBD250).

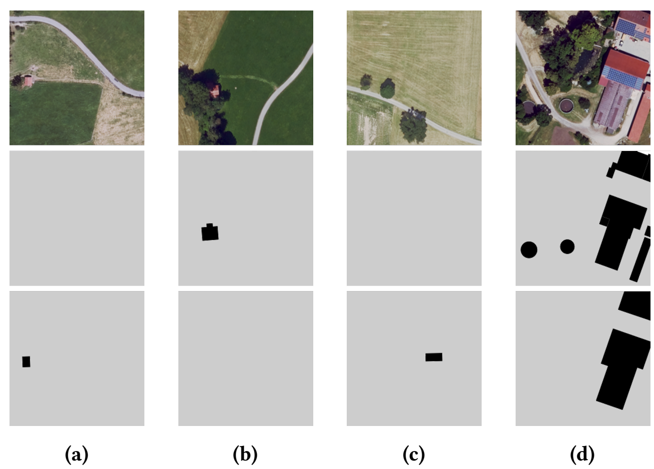

Figure 3: Examples: (a) Missing in Official, (b) Missing in OSM, (c) Missing in Orthophoto, (d) Semantic Difference (Top: Orthophoto, Middle: Official building footprints, Bottom: OSM)

4.2 Spatial Discrepancy

Another important effect one needs to consider in the BBD dataset is the diverse cases of spatial discrepancy between three datasets (i.e., DOP40, Official building footprint, and OSM). For instance, Figure 3 (a) refers to where buildings are missing in the official dataset while mapped in OSM, which implies the potential for information gain by integrating crowdsourcing information into the official dataset. Vice versa, Figure 3 (b) shows a common case of a missing building in OSM. Further, two interesting cases are demonstrated in Figure 3 (c) and (d). The former refers to a possible out-of-data OSM building geometry, and the latter shows the semantic discrepancy between official and OSM data w.r.t buildings.

5 SELECTED DATA ANALYSIS TASKS

Accurate and up-to-date geographical information on urban buildings is key for sustainable development as well as monitoring of human wellbeing. In this context, the BBD dataset paves the pathway toward important and promising data mining and image analysis tasks. As shown in Figure 4 ((a) to (d)), these tasks include but are not limited to the following:

Building Detection - As one of the preliminary analyses, a building detection task refers to a binary classification question, in which for each image, one shall only decide whether there are buildings inside or not, for example, in [7]. Note that for such a task, one may take a smaller subset of the BBD such that the distribution of patches with and without buildings is somehow balanced.

Building Segmentation - The task of building segmentation is easily defined from only image pairs by looking for algorithms that take the input DOP40 image and output the binary mask generated by rendering the spatial databases. The state-of-the-art segmentation models range from U-Net [8] to Vision Transformers [1]. An early work of building segmentation is [2].

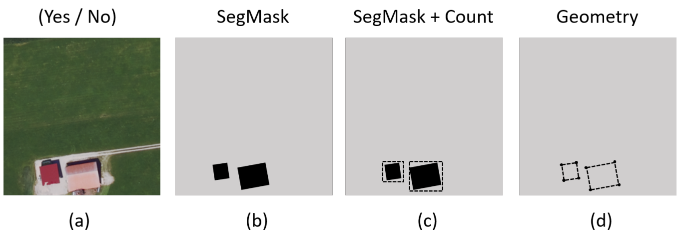

Figure 4: Exemplary analysis tasks with the BBD dataset: (a) Detection, (b) Segmentation, (C) Instance Extraction, (d) Footprint Geometry Extraction.

Building Instance Extraction - Besides the bare image segmentation, one can also extract the identity by counting the number of buildings in an image. We denote such a task as building instance extraction. For instance, in [12], Zhao et al. present an early work in this direction using mask R-CNN [6].

Building Footprint Geometry Extraction - A more interesting task could be extracting the building footprint geometry as given in the Shapefile (e.g., Official building footprints or OSM). Such a similar task was investigated using point clouds data in [9], while it remains underexplored for VHR images and GIS data.

Multimodal Data Quality Assessment - The BBD data provides a valuable resource for data quality estimation. Specifically, by comparing to the official building footprint, one can derive a list of quality metrics (e.g., completeness, position, and shape accuracy like in [3]). More importantly, the unique characterization of the BBD data enables researchers to explore and extend existing data quality estimation methods from a novel perspective of combining extrinsic and intrinsic methods.

Multimodal Localization - Cross-View Geo-Localization [11] aims to localize a ground view through matching against a reference database of overhead imagery with known locations. As shown in Figure 3, important landmarks can be missing when only relying on orthophotos, and current datasets like [13] are not investigating this challenge. Therefore, BBD enables multimodal localization and additional cues when there are discrepancies between the orthophotos and ground views.

6 CONCLUSION AND OUTLOOK

With the BBD dataset, we hope the challenges of extracting real-world geometric objects will receive greater attention in the community. These challenges are highly interdisciplinary, including aspects of computer vision, spatial computing, and Geographic Information Systems. A knowledge-driven intelligent system would rely on the ability of computer vision systems to understand the meaning and belongings of individual pixels but would also want to generate geometric representations as polygons for further processing. Semantic knowledge like rectangular shapes and parallel and orthogonal relations can optimize these polygonal representations. Hence, the step from segmented pixels to geometric footprint is far from trivial and might also need the original image information.

In addition to these foundational questions of visual information extraction, the dataset provides a massive collection of real-world data with all spatial complexities, making GeoAI significantly different from classical AI or computer vision. We are unaware of a curated dataset of comparable size, especially not of one that provides two distinct label sources: one official (i.e., building footprints) and one crowdsourced (i.e., OSM) together with the relevant geometry information as GIS files.

Moreover, the BBD dataset provides a large-scale, high-quality dataset with huge opportunities for revisiting extrinsic and intrinsic OSM data quality assessment methods from a novel perspective of integrating automatic mapping systems based on VHR image and GeoAI models.

We plan to update the dataset roughly every second year, each time when the Bavarian government releases new official geodata. Additionally, we then collect updated crowdsourced data from a similar timestamp. In future work, we plan to extend the current dataset step by step by including additional spatial object types, such as streets, water bodies, and diverse land use land cover classes (e.g., forest, commercial, industrial, and cropland).

REFERENCES

[1] Lucas Beyer, Pavel Izmailov, Alexander Kolesnikov, Mathilde Caron, Simon Kornblith, Xiaohua Zhai, Matthias Minderer, Michael Tschannen, Ibrahim Alabdulmohsin, and Filip Pavetic. 2023. FlexiViT: One Model for All Patch Sizes. arXiv:2212.08013 [cs.CV]

[2] Benjamin Bischke, Patrick Helber, Joachim Folz, Damian Borth, and Andreas Dengel. 2019. Multi-task learning for segmentation of building footprints with deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1480–1484.

[3] Hongchao Fan, Alexander Zipf, Qing Fu, and Pascal Neis. 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science* 28, 4 (2014), 700–719.

[4] Michael F. Goodchild. 2007. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* 69 (08 2007), 211–221. https://doi.org/10.1007/s10708-007-9111-y

[5] Michael F. Goodchild and Linna Li. 2012. Assuring the quality of volunteered geographic information. *Spatial Statistics* 1 (2012), 110 – 120. https://doi.org/10.1016/j.spasta.2012.03.002

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2018. Mask R-CNN. arXiv:1703.06870 [cs.CV]

[7] Hao Li, Benjamin Herfort, Wei Huang, Mohammed Zia, and Alexander Zipf. 2020. Exploration of OpenStreetMap missing built-up areas using twitter hierarchical clustering and deep learning in Mozambique. *ISPRS Journal of Photogrammetry and Remote Sensing* 166 (2020), 41–51.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI)*. Springer, 234–241.

[9] Hunsoo Song and Jinha Jung. 2022. Challenges in Building Extraction from Airborne LiDAR Data: Ground-Truth, Building Boundaries, and Evaluation Metrics. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems* (Seattle, Washington) *(SIGSPATIAL '22)*. Association for Computing Machinery, New York, NY, USA, Article 118, 4 pages. https://doi.org/10.1145/3557915.3565983

[10] Martin Werner and Hao Li. 2022. AtlasHDF: an efficient big data framework for GeoAI. In *Proceedings of the 10th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*. 1–7.

[11] Scott Workman, Richard Souvenir, and Nathan Jacobs. 2015. Wide-Area Image Geolocalization with Aerial Reference Imagery. In *IEEE International Conference on Computer Vision (ICCV)*. 1–9. https://doi.org/10.1109/ICCV.2015.451 Acceptance rate: 30.3%.

[12] Kang Zhao, Jungwon Kang, Jaewook Jung, and Gunho Sohn. 2018. Building extraction from satellite images using mask R-CNN with building boundary regularization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 247–251.

[13] Sijie Zhu, Taojiannan Yang, and Chen Chen. 2021. Vigor: Cross-view image geo-localization beyond one-to-one retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3640–3649.