BUILDING TYPE CLASSIFICATION WITH INCOMPLETE LABELS

Nikolai Skuppin [1,2], *Eike Jens Hoffmann* [1,2], *Yilei Shi* [3], *Xiao Xiang Zhu* [1,2]

[1] Data Science in Earth Observation, Technical University of Munich (TUM), Munich, Germany
[2] Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany
[3] Remote Sensing Technology, Technical University of Munich (TUM), Munich, Germany

ABSTRACT

Buildings can be distinguished by their form or function and maps of building types can be used by authorities for city planning. Training models to perform this classification requires appropriate training data. OpenStreetMap (OSM) data is globaly available and partly provides information on building types. However, this data can be incomplete or wrong. In this work a U-Net is trained to group buildings into one of the three major function classes (*commercial/industrial*, *residential* and *other*) using incomplete OSM data or ground-truth cadastral data. The model achieves overall accuracies of 72 and 75 percent. Given the OSM data has only around 20 percent of the ground truth labels this shows the incomplete data can be used to train for the building classification task.

Index Terms— Building-types, OSM, Cadastral, Semantic Segmentation, Remote-Sensing

1. INTRODUCTION

Buildings can be disinguished by their structure [1, 2, 3]. or by their function [4, 5, 6]. Such information is needed by local authorities for city planning and allows comparison of different cities. However, most of the studies focusing on building type classification develop a custom classification scheme, which applies only to a regional area, hampering global comparison. Remote-sensing-based, map-based and social-sensing based methods can be distinguished [6]. Map-based and social-sensing based methods allow for a finer classification granularity (e.g. distinguishing schools and hospitals). However, these methods cannot be transferred to other geographic areas, where this data is not available. The same issue can occur when using remote-sensing data alone, as in [4]. They describe a rule-based classification scheme, which fuses information obtained from multiple sensors. This classification scheme requires expert knowledge to formulate rules, which will likely not apply to different geographic regions.

Instead of manually designing such rules, machine learning can be used to train a model to learn classification rules based on labeled input data. Training such a model with data from different geographic areas should resolve this issue. Yet, obtaining high quality labeled data is time consuming and costly. As an alternative, crowd-sourced data like OpenStreetMap [7] can be used. However, while this data is available in abundance it is neither necessarily complete nor correct. In this paper the effect of using incomplete OSM data to train a model for building type classification is studied. Therefore, a U-Net [8] is trained with OSM data and compared against a baseline trained on official cadastral data. The workflow to generate building type maps from satellite images is given in the following and first results are presented.

2. METHODOLOGY

Building type classification from remote sensing images may be treated as a semantic segmentation problem with potentially noisy and incomplete labels.

The task can be split into two sub-problems: building detection and building classification, i.e. separating buildings from background and classifying buildings into one of the three classes: *commercial/industrial*, *residential* and *other*. The final building type maps can be obtained by merging the classification maps with the building footprints. A similar approach for land-use mapping is described in [9]. Figure 1 shows the overall workflow. The classification maps are generated from the satellite images. The building footprints are either obtained from the satellite images, or, if available, from map data such as OSM. This paper focuses on the classification and post-processing part, as there are various other studies focusing on the building extraction task [10, 11, 12].

OSM data and cadstral ground truth was obtained for ten German cities. Cadastral data is assumed to be complete, whereas, the OSM data is regarded as noisy and incomplete (missing buildings and labels). A subset of nine cities is used to train a U-Net [8], which is tested on the cadastral data of the remaining city. Since the classes are imbalanced (most pixels are background, followed by *residential*, *commercial/industrial*, and *other*) a weighted cross-entropy loss is used. The weights are obtained from the inverse class frequencies (number of pixels belonging to one class divided by the sum of pixels contributing to the loss).

For the OSM data there are many *unlabeled buildings* (see figure 2). To test if this information can help the model to learn the classification task, a combined loss function, which

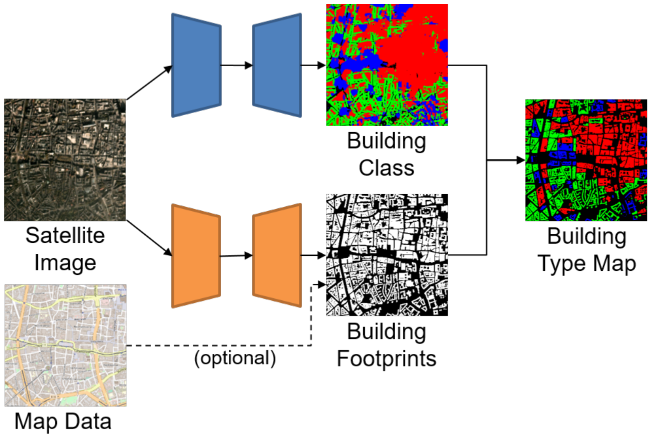


Fig. 1. Proposed workflow. Building type maps are generated from satellite images. Building footprints for post-processing are optionally generated from map data. This work focuses on the upper branch, taking the footprints as given.

consists of one building detection loss and one classification loss with corresponding weighting parameters, is used.

$$L_{tot} = \alpha_{det} L_{det} + \alpha_{class} L_{class} \qquad (1)$$

where α_{det} and α_{class} are weighting factors to adjust the losses, L_{det} and L_{class} are two cross-entropy losses. The classification loss is only calculated for the pixels belonging to a labeled built-up class (i.e. *commercial/industrial*, *residential* and *other*) . For this a softmax is applied to the model output and the values for the built-up classes are summed to obtain the building probability, which is then fed to the loss function.

The results are compared with a training using only one loss function for both the detection and classification task, where the *unlabeled buildings* are treated as background. This covers the case of incomplete building footprints in OSM, which is especially observed in other geographic regions, where OSM is not as complete as in the studied area.

3. EXPERIMENTS & RESULTS

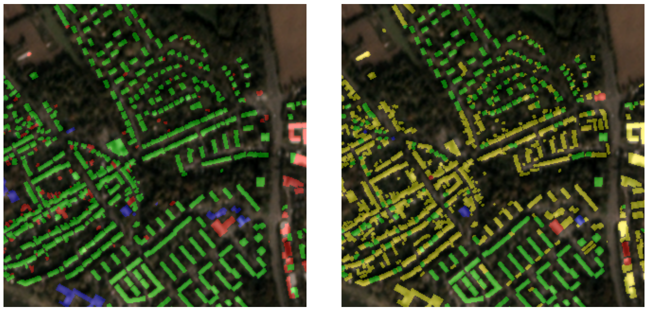
3.1. Dataset

The dataset consists of cadastral and OSM vector data for ten cities from North Rhine-Westphalia (Germany), which was obtained from [13] and [7]. The ten German cities are: *Bielefeld, Bochum, Bonn, Cologne, Dortmund, Duesseldorf, Duisburg, Essen, Muenster* and *Wuppertal*.

Building footprints from OSM together with their semantic tags *building*, *amenity*, and *shop* were extracted. The cadastral data contains building functions by default. Semantic information from both sources was then homogenized to a common labeling scheme of *commercial/industrial*, *residential*, *other*, where mixed cases were assigned to *residential* or the majority class. An additional *no label* class to handle buildings without any function tag (see Fig. 2) was added.

All vector data was then rasterized to 3 m GSD, paired with the Planet basemap images [14], and label/image patches of 320x320 pixels were cut from the dataset.

The overall accuracy between building instances in cadastral and OSM data is around 33%. This increases to around 95% when omitting buildings with *no label* in OSM data, which suggests missing OSM labels to be the main difference between both data. Fig. 2 supports this observation. For the OSM data many buildings are colored in yellow (i.e. building footprint but no class information is available). Comparing pixels with class information (red, green and blue) very few differences are observed. Therefore, the OSM data is regarded as incomplete, but not noisy (i.e. having wrong labels).



(a) Cadastral (b) OSM

Fig. 2. Overlay of image and labels for cadastral (a) and OSM (b) data. Red is commercial/industrial, green is residential, blue is other and yellow is unlabeled building. The OSM image shows many unlabeled buildings.

3.2. Metrics

Overall accuracy (OA) and mean Intersection over Union (mIoU) are calculated. For each metric three values are reported. *All* is calculated from all pixels (buildings and background), whereas *Build* is calculated considering only the labeled built-up pixels. The *Instance* metric is calculated after majority voting per building instance, which is obtained from the cadastral data.

3.3. Training

A U-Net [8] is trained on cadastral or OSM data. The optimizer is Adam with default settings in PyTorch, with a batch size of 16, an initial learning rate of 1e-3, a minimum learning rate of 1e-7, and a learning rate reduction by a factor of 10, if the validation loss did not improve for five epochs. For final model evaluation the model with the lowest validation loss is

Table 1. Mean and standard deviation of overall accuracy (OA) and mean Intersection over Union (mIoU) for models trained on cadastral (CAD) and OpenStreetMap (OSM) data. Pixel-based metrics including background detection task (All), pixel-based metrics for labeled buildings only (Build) and instance-based metrics for labeled buildings (Instance).

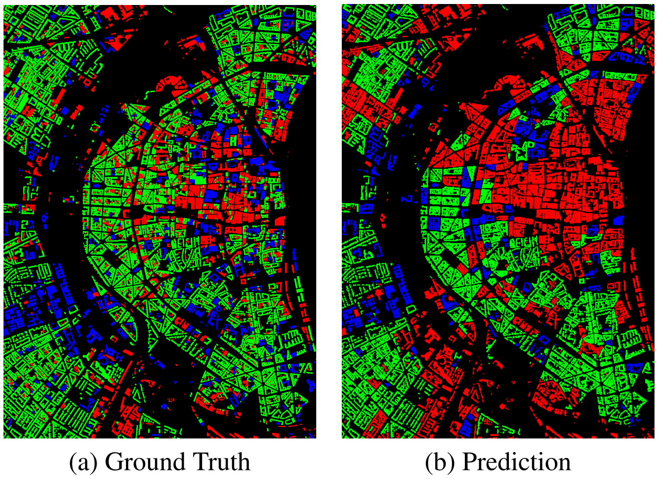
	OA			mIoU		
	All	Build	Instance	All	Build	Instance
CAD $\alpha_{det} = \alpha_{class} = 1.0$	0.56 ± 0.02	0.75 ± 0.01	0.75 ± 0.01	0.25 ± 0.01	0.51 ± 0.01	0.42 ± 0.01
CAD $\alpha_{det} = 0.0$	0.10 ± 0.00	0.74 ± 0.02	0.74 ± 0.02	0.07 ± 0.00	0.50 ± 0.01	0.40 ± 0.01
CAD $\alpha_{det} = 0.0$, L_{class}	0.69 ± 0.02	0.73 ± 0.02	0.72 ± 0.03	0.32 ± 0.01	0.50 ± 0.03	0.40 ± 0.02
OSM $\alpha_{det} = \alpha_{class} = 1.0$	0.55 ± 0.02	0.72 ± 0.01	0.72 ± 0.02	0.24 ± 0.01	0.48 ± 0.01	0.38 ± 0.01
OSM $\alpha_{det} = 0.0$	0.10 ± 0.01	0.72 ± 0.02	0.72 ± 0.03	0.07 ± 0.00	0.48 ± 0.01	0.37 ± 0.01
OSM $\alpha_{det} = 0.0$, L_{class}	0.68 ± 0.01	0.68 ± 0.03	0.68 ± 0.03	0.31 ± 0.01	0.46 ± 0.02	0.36 ± 0.02
OSM 30% labeled buildings removed	0.55 ± 0.02	0.71 ± 0.01	0.70 ± 0.01	0.24 ± 0.02	0.47 ± 0.01	0.35 ± 0.04
OSM 30% buildings removed	0.50 ± 0.09	0.68 ± 0.06	0.67 ± 0.07	0.22 ± 0.04	0.45 ± 0.05	0.37 ± 0.01

chosen. The training is performed on nine of the ten cities and all models are tested on the cadastral data from Cologne. A five fold split is used for training and testing. Samples are first sorted in east-west direction and then 80% are used for training and 20% for validation.

3.4. Results

Three settings for both the cadastral and OSM data are compared. One run with $\alpha_{det} = \alpha_{class} = 1.0$, one run with $\alpha_{det} = 0$ and another run with $\alpha_{det} = 0$ and including the background class in L_{class} (i.e. using one cross-entropy loss for both the detection and classification task). In another preliminary experiment we randomly remove 30% of the labeled buildings in OSM. We either remove only labeled buildings or also remove 30% of the unlabeled buildings. Table 1 summarizes the experiment results. The reported metrics are the mean and standard deviation of the five training runs. As expected, α_{det} has a huge impact on the metrics covering the detection task (*All*). However, when focusing only on the classification task (*Build* and *Instance*), α_{det} seems to have no impact. The combined loss (detection + classification) achieves better accuracy when considering both the detection and the classification task. Yet, it has 2 to 4 percentage points lower accuracy for the classification task alone. Comparing pixel- and instance-based metrics the overall accuracy is comparable, however, the instance-based mIoU is approximately ten percentage points lower. Overall, the cadastral training achieves 3-4% higher accuracy than the OSM training. If 30% of the labeled buildings are removed there is only a slight drop in overall accuracy and mIoU. However, when also 30% of the *unlabeled buildings* are removed the accuracy drops and the variance increases.

Figure 3 shows ground truth and prediction after post-processing, i.e. conducting majority voting for each building for a scene in Cologne. The predicted map cannot reproduce very fine structures from the ground truth. Especially, in the center the model predicts only *commercial/industrial* buildings, where there should be a mixture of all three classes.


(a) Ground Truth

(b) Prediction

Fig. 3. Ground truth and prediction after post-processing (majority voting per building instance) for Cologne. Red is commercial/industrial, green is residential and blue is other.

3.5. Discussion

The previous experiments show that both cadastral and OSM data can be used for training a network for building type classification. When focusing only on the classification task the cadastral training achieves slightly higher accuracy. Yet, this could also be caused by solely testing the model on cadastral data, giving the model trained on cadastral data a slight advantage.

Separating building detection and classification task appears to enable slightly higher classification accuracy compared to learning both tasks in one loss function. However, teaching the model to also detect buildings ($\alpha_{det} > 0$) bears no benefit for the classification metrics. This suggests that separating building detection from classification and fusing the footprints with the classification map is a valid approach for generating building type maps.

The preliminary experiment with removed labels suggests training on incomplete class labels but complete building footprints to be better than training on incomplete class labels and building footprints. Further experiments with different levels of completeness are needed.

The model is not able to reproduce the fine structure in the city center. While it is comparably easy to distinguish a huge industrial complex from a suburban residential area from remote sensing images, it is nearly impossible to distinguish an administrative from some commercial building in a dense city center. This may be improved by using an advanced post-processing or adding supplementary data.

4. CONCLUSION

This work shows that incomplete OpenStreetMap data can be used to train a building type classification model, which performs comparable to a model trained on complete ground truth cadastral data. Testing both models on cadastral data the best performing cadastral model achieves 75% overall accuracy and the best performing OSM model 72%. Given that OSM has significantly less labeled buildings these results show the feasibility of training such a model from OSM data. In the future we will extend the training to noisy OSM data (i.e. confused class labels) and test if semi- and self-supervised methods can be applied to the given problem.

5. REFERENCES

[1] Michael Wurm, Andreas Schmitt, and Hannes Taubenböck, "Building Types' Classification Using Shape-Based Features and Linear Discriminant Functions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1901–1912, 2016.

[2] Arthur Lehner and Thomas Blaschke, "A Generic Classification Scheme for Urban Structure Types," *Remote Sensing*, vol. 11, no. 2, pp. 173, Jan. 2019.

[3] Mengmeng Li, Elco Koks, Hannes Taubenböck, and Jasper van Vliet, "Continental-scale mapping and analysis of 3D building structure," *Remote Sensing of Environment*, vol. 245, pp. 111859, Aug. 2020.

[4] Tanakorn Sritarapipat and Wataru Takeuchi, "Building classification in Yangon City, Myanmar using Stereo GeoEye images, Landsat image and night-time light data," *Remote Sensing Applications: Society and Environment*, vol. 6, pp. 46–51, Apr. 2017.

[5] Jionghua Wang, Haowen Luo, Wenyu Li, and Bo Huang, "Building Function Mapping Using Multisource Geospatial Big Data: A Case Study in Shenzhen, China," *Remote Sensing*, vol. 13, no. 23, pp. 4751, Jan. 2021, Number: 23 Publisher: Multidisciplinary Digital Publishing Institute.

[6] Wei Chen, Yuyu Zhou, Qiusheng Wu, Gang Chen, Xin Huang, and Bailang Yu, "Urban Building Type Mapping Using Geospatial Data: A Case Study of Beijing, China," *Remote Sensing*, vol. 12, no. 17, pp. 2805, Jan. 2020, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

[7] "Openstreetmap," `https://www.openstreetmap.org`.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, Eds., Cham, 2015, Lecture Notes in Computer Science, pp. 234–241, Springer International Publishing.

[9] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huanfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sensing of Environment*, vol. 237, pp. 111322, Feb. 2020.

[10] Ksenia Bittner, Fathalrahman Adam, Shiyong Cui, Marco Körner, and Peter Reinartz, "Building footprint extraction from vhr remote sensing images combined with normalized dsms using fused fully convolutional networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2615–2629, 2018.

[11] Yilei Shi, Qingyu Li, and Xiao Xiang Zhu, "Building segmentation through a gated graph convolutional neural network with deep structured feature embedding," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 184–197, Jan. 2020.

[12] Haonan Guo, Xin Su, Shengkun Tang, Bo Du, and Liangpei Zhang, "Scale-robust deep-supervision network for mapping building footprints from high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10091–10100, 2021.

[13] "GEOportal.NRW," `https://www.geoportal.nrw`, Accessed: 2021-08-05.

[14] "Planetscope," `https://www.planet.com`.