MULTI-SCALE BUILDING MAPS FROM AERIAL IMAGERY

Y. Feng[1,*], C.Yang[2], M. Sester[1]

[1] Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany - (feng, sester)@ikg.uni-hannover.de
[2] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany - yang@ipi.uni-hannover.de

Commission III, WG III/1

KEY WORDS: Multi-scale Building Map, Cartographic Generalization, Aerial Imagery, Multiple Representations

ABSTRACT:

Nowadays, the extraction of buildings from aerial imagery is mainly done through deep convolutional neural networks (DCNNs). Buildings are predicted as binary pixel masks and then regularized to polygons. Restricted by nearby occlusions (such as trees), building eaves, and sometimes imperfect imagery data, these results can hardly be used to generate detailed building footprints comparable to authoritative data. Therefore, most products can only be used for mapping at smaller map scale. The level of detail that should be retained is normally determined by the scale parameter in the regularization algorithm. However, this scale information has been already defined in cartography. From existing maps of different scales, neural network can be used to learn such scale information implicitly. The network can perform generalization directly on the mask output and generate multi-scale building maps at once.
In this work, a pipeline method is proposed, which can generate multi-scale building maps from aerial imagery directly. We used a land cover classification model to provide the building blobs. With the models pre-trained for cartographic building generalization, blobs were generalized to three target map scales, 1:10,000, 1:15,000, and 1:25,000. After post-processing with vectorization and regularization, multi-scale building maps were generated and then compared with existing authoritative building data qualitatively and quantitatively. In addition, change detection was performed and suggestions for unmapped buildings could be provided at a desired map scale.

1. INTRODUCTION

Multi-scale topographic map is one of the important surveying products from National Mapping Agencies (NMAs). Buildings are the most basic elements in topographic maps. There has been an increasing interest in automatic buildings extraction from aerial imagery. Many studies have been conducted to extract buildings by image segmentation. The outputs of current models are mostly binary pixel masks, where buildings' locations or boundaries are encoded. They can hardly be used directly to generate map products. Post-processing is needed to generate building in vectors. Meanwhile, there are only very few studies on end-to-end deep models predicting building in a vector representation, e.g. Li et al. (2019).

Two main issues can be identified for the outputs of building masks. The shapes of the detected buildings are often not clean, which may contain noisy or blurred pixels at building boundaries. This makes post-processing less stable. Another issue is scale, i.e. to what extent the details of the blobs should be preserved. Many post-processing techniques were applied to regularize the building outlines to vectors, such as Douglas-Peucker (He et al., 2019), or a combination of local and global regularization (Xie et al., 2018). However, they often require a scale parameter, which defines the level of detail to preserve. Visual inspection or trial-and-error are popular approaches to determine the optimal scale parameters (Chen et al., 2018).

The scale of the generated buildings determines its application scenarios. Freire et al. (2014) compared the extracted buildings from QuickBird multispectral data (2.4 m resolution) with the authoritative map at 1:1,000, 1:5,000, and 1:10,000. The results show many extracted buildings could meet the topographic standards for scale 1:10,000 but hardly any for scale 1:1000. Obviously, the resolution of the images affects which scale can be derived. Higher-resolution imagery can be used to derive more detailed building shapes. However, restricted by the imagery quality, building eaves and occlusions (e.g. nearby objects like trees), these results can still hardly be used to generate very detailed building footprints (e.g. 1:1,000 or 1:5,000).

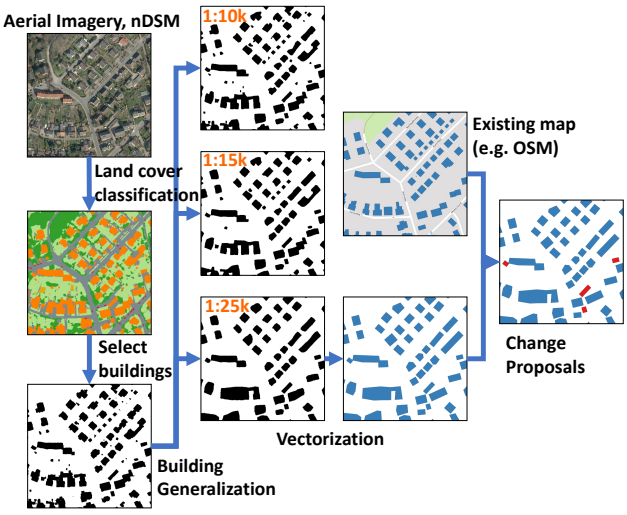

Figure 1. Workflow of the proposed method.

The scale dependent representation of objects in maps is defined in cartography and achieved via different generalization operations; also, existing maps can be used to study and learn such

* Corresponding author

scale-dependent information. Thus, instead of choosing this parameter by trial-and-error, buildings can be generated according the scale learned from existing maps. In our previous work (Sester et al., 2018; Feng et al., 2019), a deep convolutional neural network (DCNN) was successfully applied for the cartographic generalization task. By converting the building vectors to binary raster maps, the network can perform generalization for building patterns at three map scales. The learned models comprise the generalization operations aggregation, simplification and elimination. In this way, very small buildings are eliminated, very close buildings are combined, and complex shapes are simplified for the map at higher scale level. The outlines, however, still contain minor irregularities, which can be refined in a post-processing step.

The idea of this paper is to apply the pre-trained building generalization models to the blobs identified by building detection models. Our model can regulate the blobs' shape and unify the scales of individual buildings. With this, multi-scale building maps can be produced directly from remote sensing observations. This paper details the three components of the proposed pipeline method in Section 2. The model output was evaluated qualitatively and quantitatively by a comparison with authoritative building data in Section 3. Change detection was performed by comparing it with OpenStreetMap data. A conclusion and an outlook are given in the last section.

2. METHOD

In this work, a pipeline method is proposed to generate building maps at three map scales directly from aerial imagery and a nDSM. The workflow of this pipeline method is visualized in Figure 1. It consists of three components, a land cover classification network for building detection, a cartographic generalization network and a post-processing step for regularizing building shapes. In addition, we can perform a change detection compared to an existing map and make suggestions for the unmapped buildings at a desired map scale.

2.1 Building detection from land cover classification

Our DCNN for land cover classification, referred to as FuseEnc (Yang et al., 2019), is based on SegNet (Badrinarayanan et al., 2017), requiring input patches of 256 x 256 pixels. Like SegNet, FuseEnc applies a symmetric encoder-decoder structure. There are four convolution blocks in the encoder part, each consisting of three convolutional layers followed by batch normalization (BN) (Ioffe et al., 2015) and a rectified linear unit (ReLU) for non-linearity. At the end of the block, there is a max-pooling layer. FuseEnc requires two inputs, resulting in a two-branch encoder. The features of each branches are fused at the end of encoder. Symmetrically, the decoder part consists of four convolution blocks, each starting with an upsampling layer via bilinear interpolation. Three convolution layers, BN and ReLU are followed. For all convolutional layers, the filter size is 3 x 3. We apply the learnable skip-connections introduced by (Yang et al., 2019) to connect feature maps from encoder to the corresponded decoder part. Figure 2 shows the network structure.

All parameters of convolutional layers are learned during in the training process, which is based on stochastic mini-batch gradient descent (SGD) using backpropagation for computing the gradients. As objective loss function, the extended focal loss

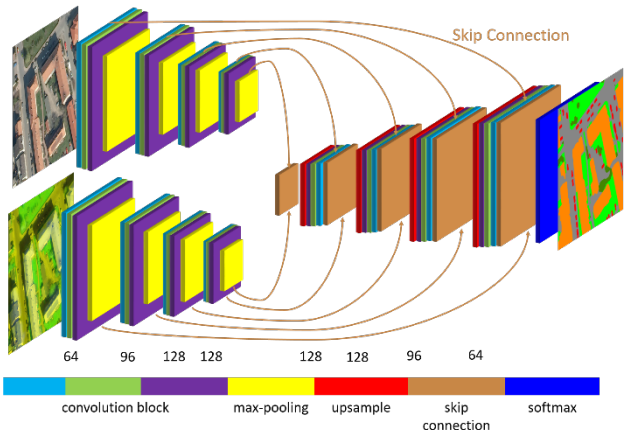

Figure 2. The architecture of FuseEnc. The numbers indicate the number of filters in the corresponding block.

(Yang et al., 2019) is applied. In the training procedure, we applied weight decay with 0.0005, a step learning policy and used a mini-batch size of 4. The learning rate was set to 0.01 and decreased to 0.001 after 30 epochs in a total of 50 epochs training. The implementation of FuseEnc is based on tensorflow[1].

FuseEnc was evaluated using the city of Hameln (Germany), covering an area of 2 km x 6 km. There are digital orthophotos (DOP), a DTM, a DSM derived by image matching available. The DOP are multispectral images (RGB + infrared / IR) with a GSD of 20 cm. We generated a normalized DSM (nDSM) by subtracting the DTM from DSM. For FuseEnc, the first branch uses RGB data as input and the another branch uses composite images as input, consisting of bands of Red, Infrared and nDSM. The reference for land cover consist of 37 manually labelled image patches, each covering 1000 x 1000 pixels (200 m x 200 m). For these datasets, we distinguish 8 land cover classes: *building*, *sealed area*, *bare soil*, *grass*, *tree*, *water*, *car* and *others*. In this work, we are focusing on the results of the building segmentation.

In the tests, we split each image into four non-overlapping tiles of size 500 x 500 pixels, resulting in 148 tiles. These tiles are randomly split into three groups of equal size for three-fold cross validation. Each tile is split into four overlapping patches corresponding to the input size of the CNN (256 x 256 pixels). In each test run, one group of tiles is used for testing and the others are used for training. Finally, we report the average overall accuracy (OA) and average F1 score over the three groups. In training, we applied data augmentation by flipping all training patches in horizontal and vertical directions and by applying rotations of 90, 180 and 270 to all patches. In order to classify the whole area of Hameln, we use the trained weights of the first group in the cross validation procedure. In the end, we obtained OA of 89.1% and average F1 score of 81.8% over eight land cover classes. For class building, its F1 score is 94.4%.

2.2 Cartographic generalization for building masks

Our initial attempt in Sester et al. (2018) demonstrated that DCNNs are able to learn multiple cartographic generalization operations for buildings in one single model in an implicit way. DCNN architectures U-net (Ronneberger et al., 2015), residual U-net (Zhang et al., 2018), and Generative Adversarial Network

[1] https://www.tensorflow.org/

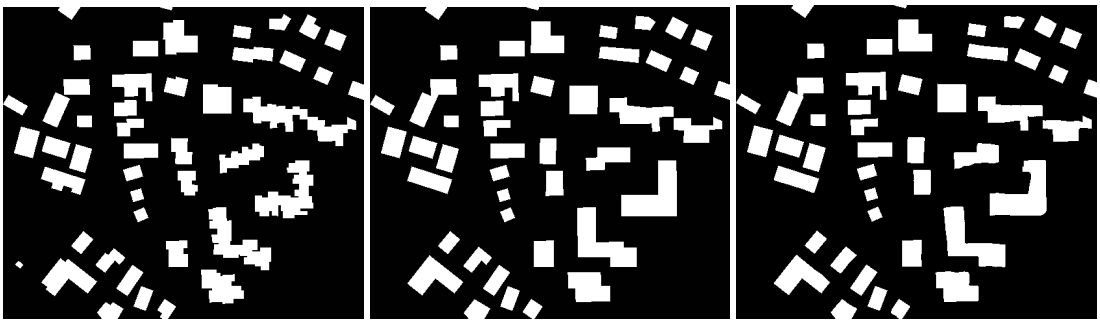

Figure 3. Example of the input (left), reference (middle) and output (right) of the building generalization network at map scale 1:15,000

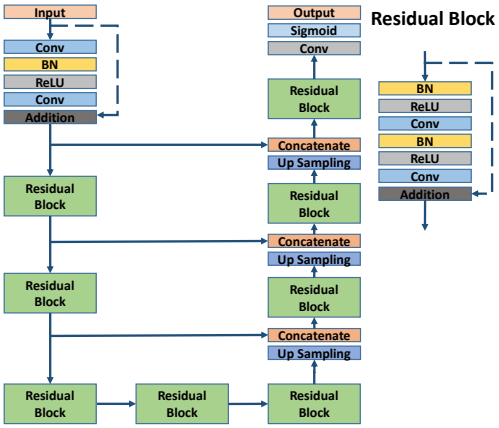

Figure 4. Residual U-net for building generalization.

(GAN) (Goodfellow et al., 2014) were further tested and evaluated for the building generalization task in Feng et al. (2019). The result shows that residual U-net is a better solution for this task compared with the other two network architectures. Figure 4 shows the network structure of residual U-net. It is an extension of U-net architecture, where the convolutional layers in U-net are replaced by residual unit (He et al., 2016), which consists of batch normalization (BN), ReLU (Rectified Linear Unit) activation and one convolutional layer (as shown in Figure 4 upper right).

OpenStreetMap (OSM)[2] building polygons, in the area of Stuttgart, Germany, were used as input data for training this network. It had an approximate scale of 1:5,000. The building polygons were generalized using software CHANGE (Powitz, 1993) into three target map scales, namely 1:10,000, 1:15,000, and 1:25,000. The target scale determined different parameters, such as the minimum length of a facade element (3 m, 4.5 m, 7.5 m) or the minimum area (9, 20, 56 m^2) to be preserved. The building polygons with and without generalization were rasterized in 0.5 m $\times$ 0.5 m grids. At this grid size the details of the building ground are ensured to be preserved during the rasterization process. The rasterized binary images are the same size 42,800 pixel $\times$ 35,000 pixel. For each map scale, correspondent input and output images pairs were generated by partitioning the raster maps into tiles (128 pixel $\times$ 128 pixel) without overlaps. 31,760 tiles were used for training and 3528 tiles (about 10%) were used for validation.

The model was evaluated for an area of 1.2 km $\times$ 1.2 km outside the training data region. Our model can achieve an pixel-wise error rate of 0.32%, 0.49%, 1.16% for map scales 1:10,000, 1:15,000, and 1:25,000 respectively. A qualitative analysis indicates that the networks have learned the simplification of the buildings for the respective scales: small buildings are eliminated in smaller scales; close, neighboring buildings are aggregated, and outlines are simplified by eliminating small extrusions and indentations. In Figure 3, an example is presented, where also a very complex shape of the buildings on the left are generalized properly to the results on the right at map scale 1:15,000. Visually, it is close to the reference in the middle.

[2] https://www.openstreetmap.org

2.3 Post-processing

Since the building blobs were generalized according to a target map scale, these buildings can then be reconstructed simply with the existing methods for building vectorization and regularization. The Marching Cubes algorithm (Lorensen and Cline, 1987) was used to extract polygons from the generalized results. In order to regularize the polygons to structured building shapes, we used the polylines simplification method proposed by Gribov (2017) to reduce the unnecessary vertices and preserve the shape of polylines. This method was realized in the *Regularize Building Footprint* module provided in arcpy library from ArcGIS 10.4 (ESRI, 2018). In order to preserve the original angles of the polygon edges, the method *ANY_ANGLE* and tolerance of 2m are used as input parameters.

Figure 5. Regularization of building footprints by completing corners (left) and improving parallelism of opposite edges (right). Input vector in red and regularized vector in blue.

It was observed that extrusions and indentations are rarely preserved after applying the generalization model. The obvious problems are that many buildings have incomplete corners and do not retain the parallelism of the opposite edges. We therefore followed the ideas from Sester and Neidhart (2008) and Sester (2005), to force opposite edges to be parallel and complete the missing part at the corner of the building. The algorithm searches for edges or groups of edges which could potentially

Figure 6. Building blobs from land cover classification (upper left) and generated building vectors at map scale 1:10,000 (upper right), 1:15,000 (lower left), and 1:25,000 (lower right)

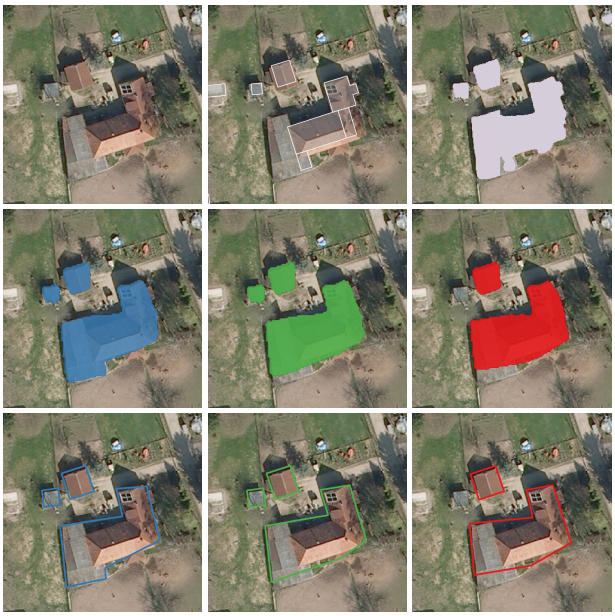

Figure 7. Building in RGB image, as footprint from ALKIS, and as blob from land cover classification (top row from left to right). Generalization outputs from the building blob at map scale 1:10,000, 1:15,000, and 1:25,000 (middle row from left to right). Post-processed building vectors at the corresponding map scales (bottom row from left to right).

build 90° or 180° degree angles by applying an angle threshold of 10°. It also checks the combination of all the edges in each polygon, which can potentially build a parallel relationship by applying a same 10° threshold. The effect of this algorithm can be observed in the two cases shown in Figure 5, where vertices are moved or removed to regularize the building polygons.

3. EXPERIMENTS AND RESULTS

In the previous section, the evaluations of the first two components, namely the land cover classification model and building generalization model, were reported separately. The evaluation of the proposed pipeline in the following will focus on the visual inspection of the generated polygon vectors and the qualitative comparison between the generated polygons and existing authoritative building footprint data.

The land cover classification model was applied to the entire Hameln (Germany) dataset covering an area of 2 km × 6 km. The input data included digital orthophotos (DOP), DTM, and DSM. The resolution of the data (0.2m) ensures that the desired scales can be derived; in fact, for the training of the generalization network a resolution of 0.5m was used, therefore, the predicted building blob map was re-sampled from 0.2 m × 0.2 m into 0.5 m × 0.5 m. The most detailed target map scale in this work is 1:10,000, where the minimum length of a facade element is 3 m. Therefore, this re-sampling step should not significantly influence the final results of this pipeline.

The building masks were then generalized to the three target map scales with the model pre-trained on the Stuttgart OSM dataset. After post-processing, the generated data were first inspected visually in section 3.1 and then compared to the building layer from German Authoritative Real Estate Cadastre Information System (ALKIS) quantitatively in section 3.2. Please note, however, that the scale of ALKIS considerably higher, approx. 1:1.000. In the end, change detection for building maps was performed by comparing the result to the OSM building polygons in section 3.3.

3.1 Multiscale building maps

Firstly, one single building is shown in Figure 7. The detected building blob from land cover classification has a complicated

boundary and small indentations can be observed. After applying the generalization model for the raster output, the boundary of the building blobs are smoother and simpler with respect to the map scale (middle row). Extrusions and indentations are nicely handled by the generalization model. Elimination of the small object in the upper right region can be observed at the map scale 1:25,000. With the post-processing step, the edges of the building are regularized and the orthogonal corners as the square-shaped building can be regulated although the building blob is not perfect. Compared with the building vectors from ALKIS (upper row in the middle), a part of the building is detected but not mapped in the authoritative data - obviously, the ALKIS is not up-to-date for this building.

In Figure 6, the building blobs are visualized, together with generalized building vectors at three target map scales, for a residential area. Simplification and aggregation of buildings can be observed with respect to the different map scales. Elimination of small buildings can be observed at smaller map scales. With the results presented above, visually, the pipeline succeeds to achieve the automatic generation of multi-scale building map from aerial images.

3.2 Quantitative Evaluation

To evaluate the proposed pipeline method quantitatively, our outputs were compared with the authoritative building footprint data - ALKIS - for the entire Hameln area. In order to compare at the same map scale, we processed the ALKIS footprints with the building generalization software CHANGE (Powitz, 1993) and generalized building footprints to the same three target map scales. IoU (Intersection over Union) was used as the metric for evaluation. In this case, the areas of the intersection and union between two building vector layers were calculated with the QGIS geoprocessing tools. The comparison between our output and generalized authoritative data at each map scale is summarized as column "IoU - original" in Table 1. An IoU score around 67% can be achieved for all map scales, which indicates a significant difference between there two respective building layers.

	IoU		IoU changes ignored	
	original	buffer 1 m	original	buffer 1 m
1:25k	66.95%	72.74%	69.60%	74.99%
1:15k	66.92%	73.62%	70.42%	75.15%
1:10k	67.29%	74.18%	70.57%	78.33%

Table 1. Quantitative evaluation of the proposed pipeline on CHANGE generalized ALKIS building vectors

However, there are several differences between our prediction and the existing map data. Over time, many changes can occur. Some of the detected buildings were not mapped in the authoritative data. Similarly, missing buildings, which were built later than the image acquisition, should not be considered as failure cases. Therefore, the buildings, which cannot find any overlaps in the other building layer, are ignored. The IoU scores in this case is summarized as the column "IoU changes ignored - original" in Table 1. It is discovered that 1.53% - 4.15% of the performance loss is due to the actual changes between these two building layers.

In addition, we discovered that a significant enlargement of the building vectors often occurs in our output. As shown in Figure 8, roofs may lead to an enlargement of buildings in all directions, and also imperfect orthophotos may lead to an enlargement of buildings in some directions. For this reason, our outputs were compared with the generalized building footprints enlarged with a buffer of 1 meter. For both cases with and without considering existing changes, the IoU scores are calculated as the columns "IoU - buffer 1 m" and "IoU change ignored - buffer 1 m" in Table 1. We can observe a significant IoU loss from 4.73% to 7.76% due to this reason. It has to be noted, that the 1m-buffer is only a rough estimate and would have to be adjusted in an object-specific way for an in depth analysis.

Figure 8. Comparison between aerial building detection and authoritative building footprint data. Differences caused by roofs lead to an enlargement of building in all directions (left), and imperfect orthophotos lead to an enlargement of buildings in some directions (right). (Buildings in white outlines from ALKIS, in orange mask from land cover classification, and in red outlines from generated building at map scale 1:10,000)

This results in general differ from the performance that the land cover classification model can achieve (see above: a classification accuracy of 94,4% was achieved). However, that is because the labels for training and evaluation the model are manually annotated based on the aerial images (i.e. for the roofs) and according to the visual appearance. In contrast, the authoritative building footprints are mostly measured by terrestrial surveying methods (i.e. capturing the walls). This essential difference should not be ignored.

3.3 Change Detection

In addition to generating building maps, this approach can perform change detection for buildings. Based on an existing map, the generated building vectors can detect the unmapped buildings and directly provide suggestions at a desired map scale. For example, in Figure 9, the building vectors from OSM in Hameln were compared with our generated building vectors at map scale 1:25,000. Buildings in orange are the generated vectors which are confirmed by OSM. Buildings in red are the proposed building vectors for the unmapped buildings in the building vector layer of OSM.

Not only could missing buildings be detected, on the contrary, we could also detect buildings that were built later than the image acquisition. In Figure 10, buildings in red are from OSM. They do not yet exist in the aerial image on left, however, they are already mapped in the current OSM building layer.

4. CONCLUSION AND OUTLOOK

In this work, a pipeline method is proposed to generate building polygons at multiple map scales directly from aerial images and nDSMs. Buildings are detected as masks using land cover classification network and then generalized into three target map scales with building generalization networks. After post-processing, our framework can provide multi-scale building maps in vector representation.

Figure 9. Comparison of the extracted building vectors at map scale 1:25,000 with OSM, generated buildings confirmed by OSM (orange) and suggestions for the unmapped buildings (red).

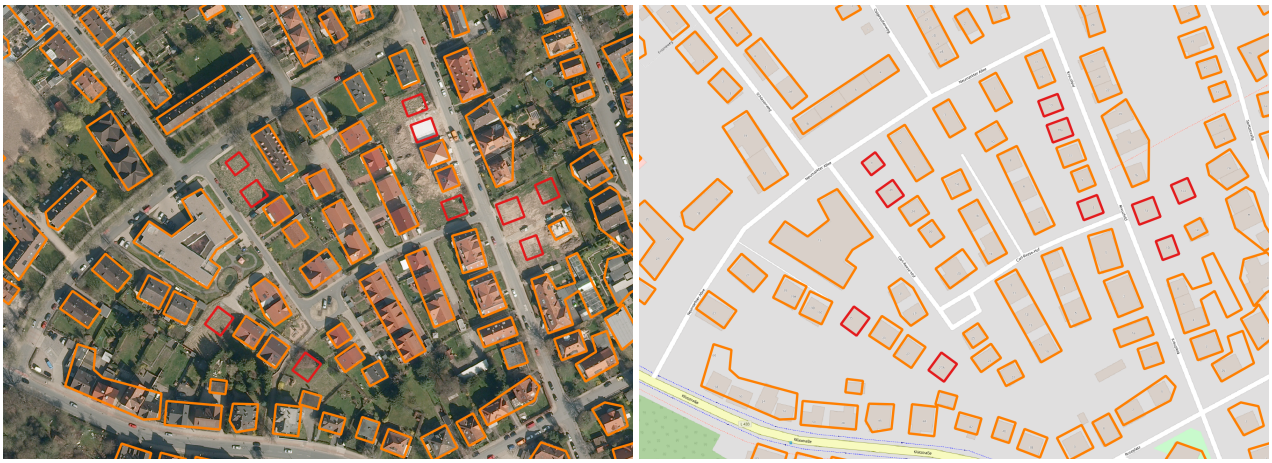

Figure 10. Comparison of the extracted building vectors at map scale 1:25,000 (orange) with OSM, OSM buildings (red) which were built later than the image acquisition were detected.

As future work, learning the generalization of other map features such as line features (road, river), free-shape polygonal features (lakes, parks) and the interplay between the map features can be further investigated. It has a great potential to facilitate the current map production and map update process.

ACKNOWLEDGEMENTS

We thank the Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN), the Landesamt für Vermessung und Geoinformation Schleswig Holstein (LVermGeo) and Landesamt für innere Verwaltung Mecklenburg-Vorpommern (LaiV-MV) for providing the test data and for their support of this project. C. Yang is an associate member of the Research Training Group i.c.sens (GRK 2159), funded by the German Research Foundation (DFG). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of a GeForce Titan X GPU used for this research.

References

Badrinarayanan, V., Kendall, A. and Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39(12), pp. 2481–2495.

Chen, G., Weng, Q., Hay, G. J. and He, Y., 2018. Geographic object-based image analysis (geobia): emerging trends and future opportunities. *GIScience & Remote Sensing* 55(2), pp. 159–182.

ESRI, 2018. Regularize Building Footprint - ArcGIS for Desktop. `https://desktop.arcgis.com/en/arcmap/10.4/tools/3d-analyst-toolbox/regularize-building-footprint.htm`. (Accessed 28.04.2020).

Feng, Y., Thiemann, F. and Sester, M., 2019. Learning cartographic building generalization with deep convolutional neural networks. *ISPRS International Journal of Geo-Information* 8(6), pp. 258.

Freire, S., Santos, T., Navarro, A., Soares, F., Silva, J., Afonso, N., Fonseca, A. and Tenedório, J., 2014. Introducing mapping standards in the quality assessment of buildings extracted from very high resolution satellite imagery. *ISPRS journal of photogrammetry and remote sensing* 90, pp. 1–9.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In: *Advances in neural information processing systems*, pp. 2672–2680.

Gribov, A., 2017. Searching for a compressed polyline with a minimum number of vertices. In: *2017 14th IAPR Inter-*

national Conference on Document Analysis and Recognition (ICDAR), Vol. 2, IEEE, pp. 13–14.

He, H., Zhou, J., Chen, M., Chen, T., Li, D. and Cheng, P., 2019. Building extraction from uav images jointly using 6d-slic and multiscale siamese convolutional networks. *Remote Sensing* 11(9), pp. 1040.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Identity mappings in deep residual networks. In: *European conference on computer vision*, Springer, pp. 630–645.

Ioffe, S., Szegedy, C. and Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*, pp. 448–456.

Li, Z., Wegner, J. D. and Lucchi, A., 2019. Topological map extraction from overhead images. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1715–1724.

Lorensen, W. E. and Cline, H. E., 1987. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics* 21(4), pp. 163–169.

Powitz, B., 1993. Computer-assisted generalization-an important software tool in gis. *International Archives of Photogrammetry and Remote Sensing* 29, pp. 664–664.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.

Sester, M., 2005. Optimization approaches for generalization and data abstraction. *International Journal of Geographical Information Science* 19(8-9), pp. 871–897.

Sester, M. and Neidhart, H., 2008. Reconstruction of building ground plans from laser scanner data. *Proceedings of the AGILE Conference, 2008* p. 111.

Sester, M., Feng, Y. and Thiemann, F., 2018. Building generalization using deep learning. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4 (2018)* 42, pp. 565–572.

Xie, L., Zhu, Q., Hu, H., Wu, B., Li, Y., Zhang, Y. and Zhong, R., 2018. Hierarchical regularization of building boundaries in noisy aerial laser scanning and photogrammetric point clouds. *Remote Sensing* 10(12), pp. 1996.

Yang, C., Rottensteiner, F. and Heipke, C., 2019. Towards better classification of land cover and land use based on convolutional neural networks. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4213, pp. 139–146.

Zhang, Z., Liu, Q. and Wang, Y., 2018. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15(5), pp. 749–753.