

DETECTION OF UNDOCUMENTED BUILDINGS USING CONVOLUTIONAL NEURAL NETWORK AND OFFICIAL GEODATA

Qingyu Li^{1,2}, Yilei Shi³, Stefan Auer¹, Robert Roschlau⁴, Karin Möst⁴, Michael Schmitt², Xiao Xiang Zhu^{1,2,*}

¹Remote Sensing Technology Institute, German Aerospace Center, Oberpfaffenhofen, Germany - (qingyu.li, stefan.auer, xiaoxiang.zhu)@dlr.de

²Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany - m.schmitt@tum.de

³Remote Sensing Technology, Technical University of Munich, Munich, Germany - yilei.shi@tum.de

⁴Bavarian Agency for Digitization, High-Speed Internet and Surveying, Munich, Germany - (robert.roschlaub, karin.moest)@ldbv.de

Commission II, WG II/6

KEY WORDS: Building Detection, Convolutional Neural Network, Deep Learning, Semantic Segmentation, Decision Fusion

ABSTRACT:

Undocumented buildings are buildings which were built years ago, but were never recorded in official digital cadastral maps. Detection of undocumented buildings is of great importance for urban planning and monitoring. The state of Bavaria, Germany, pursues this task based on high resolution optical data and digital surface models, using semi-automatic detection methods, which suffer from a high false alarm rate. In order to study the influence of sampling strategies on the performance of building detection, we have firstly designed a transferability analysis experiment, which has not been adequately addressed in the current literature. In this experiment, we test whether the trained model from a district contains valuable information for building detection in a different district. It was found that the large-scale building detection results can be considerably improved when training samples are collected from different districts. Based on the building detection results, we propose a novel framework for the detection of undocumented buildings using Convolutional Neural Network (CNN) and official geodata. More specifically, buildings are identified as undocumented, when their pixels in the output of the CNN are predicted as "building", whereas they belong to the "non-building" in the Digital Cadastral Map (DFK). The detected undocumented building pixels are subsequently divided into the class of old or new undocumented building with the aid of a Temporal Digital Surface Model (tDSM) in the stage of decision fusion. By doing so, a seamless map of undocumented buildings is generated for 1/4th of the state of Bavaria, Germany at a spatial resolution of 0.4 m, which has demonstrated the use of CNN for the robust detection of undocumented buildings at large-scale.

1. INTRODUCTION

Buildings are important geospatial targets, which indicates human settlement areas. The creation and maintenance of databases of building models have numerous applications, which involve urban planning and monitoring, Three Dimension (3D) city modeling as well as estate management. However, there are some buildings, which were built years ago, but were never recorded via terrestrial surveying and are thus missing in the real estate cadastral maps. These buildings are named as undocumented buildings, and the collection of these buildings is necessary to continue and complete these databases.

An accurate building model can be acquired through a field survey, and is documented in the Digital Cadastral Map (DFK), which is a two-dimension ground plan of buildings. However, this may require a great amount of time and manual work for large-scale mapping. Compared with traditional land survey methods, the technologies of airborne imaging and laser scanning show great potential in the task of building detection. The high-resolution airborne datasets are usually taken over large areas on the ground, which makes it convenient for detailed analysis.

The currently utilized strategy for the detection of undocumented buildings in the state of Bavaria, Germany updates building models by heuristic methods (Geßler et al., 2019) (Roschlau et al., n.d.). Firstly, from the DFK where the building ground plans are included, the Red Green Blue (RGB) color values of all

pixels of corresponding buildings in orthophoto with RGB bands (TrueDOP) are collected as a reference. There is neither tilting of elevated objects nor geometric distortion effects of the roof tops in TrueDOP. Then, the frequencies and distributions in the RGB color cube are counted to separate buildings from vegetation by an empirically selected threshold value. Additionally, with the aid of a Normalized Digital Surface Model (nDSM), the misclassifications between building and other impervious objects such as roads can be avoided with an empirically determined height threshold. However, since the heuristic definition of threshold values is not standardized, e.g. it has to be determined individually for each photo flight project, which easily leads to bias and poor generalization. Moreover, as the RGB color cubes considered for determining buildings are wide, which also involve vegetation, many trees are misclassified as buildings.

Recently, Deep Learning (DL) methods such as Convolutional Neural Network (CNN) are favoured in the remote sensing community (Zhu et al., 2017), also for the task of building detection from remote sensing data (Shi et al., 2020). This is due to their superiority in generalization and accuracy without hand-crafted features. For the CNN, the amount of training data could be reduced if the trained models that use samples created for some areas could be implemented to building detection in another different area, a process that is referred to as transferability (Li et al., 2020). However, due to the limited size and quality of existing public dataset (Vargas-Muñoz et al., 2019), the transferability can not be well investigated. Therefore, in this paper, three main contributions are made:

* Corresponding author

(1) In order to offer useful sampling strategies for similar large-scale building detection tasks, we have investigated the transferability issue further by using reference data of selected districts across the state of Bavaria, Germany and employing the CNN model and official geodata. It should be noted that the utilized official geodata in this research are with really high quality and cover the study area for 1/4th of the state of Bavaria, Germany. Therefore, this work is in an advanced position to study the transferability of trained models in large-scale building detection tasks.

(2) In this regard, we have proposed a framework for the detection of undocumented buildings, which has integrated the state-of-the-art CNN model and fully harnessed official geodata. The proposed framework can identify undocumented buildings, and meanwhile can distinguish old undocumented buildings from new undocumented buildings according to their year of construction. Firstly, a CNN model is implemented for semantic segmentation of combined nDSM and TrueDOP data into a map of "building" and "non-building" pixels. This binary map is then used to identify undocumented buildings by automatic comparison with the DFK. In order to separate old and new undocumented buildings, we select thresholds on Temporal Digital Surface Model (tDSM), which is the difference between Digital Surface Model (DSM) in two temporals.

(3) Using a single optimal CNN model derived from the training data, which are collected from 14 surveying districts, a seamless map of undocumented buildings is generated for 1/4th of the state of Bavaria, Germany at a spatial resolution of 0.4 m. The achieved results demonstrate the use of CNN for the robust detection of undocumented buildings at large-scale.

The remainder of this paper is structured as follows: The study area and official geodata considered in this research are introduced in Section 2. The CNN architecture and proposed framework for the detection of undocumented buildings in Section 3, and the experiment setup are described in Section 4. The results and discussion are presented in Section 5 and Section 6, respectively. Finally, this work is summarized and concluded in Section 7.

2. STUDY AREA AND OFFICIAL GEODATA

In this research, the study sites cover 1/4th of the state of Bavaria, Germany (see Figure 1), which include 15 districts: Ansbach, Bad Toelz, Wolfratshausen, Kulmbach, Kronach, Landau, Degendorf, Landshut, Muenchen, Regensburg, Hemau, Rosenheim, Wasserburg, Schweinfurt, and Weilheim.

There are four types of official geodata used in this study: DFK, tDSM, nDSM, and TrueDOP. DFK is the cadastral 2D ground plan of the state of Bavaria, Germany including the outline of buildings. It is provided as a vector file and based on a terrestrial surveying in the field with accuracy in the range of cm. The nDSM is a difference model between a current DSM at time point 2 (year 2017) and the Digital Terrain Model (DTM) of the scene. The nDSM highlights elevated objects above the ground such as buildings and trees. The DTM is obtained from airborne laser scanning, which is derived as a regular point grid. The DSM is obtained from an image-based surface model with a dense matching method. The tDSM is the difference of two DSMs captured at two time points, i.e., time point 1 (year 2014) and 2 (year 2017). The TrueDOP is an orthophoto with RGB bands where ortho projection and geo-localization has been

realized based on the DSM. TrueDOP is also acquired in time point 2 (year 2017). The TrueDOP represents all buildings and tall objects in the correct position without geometric distortion effects. Each district is covered by many tiles of TrueDOP, nDSM and tDSM, and each tile has a size of 2500×2500 pixel at a spatial resolution of 0.4 m.

3. METHODOLOGY

3.1 CNN Architecture

For the airborne data, a large urban area may be covered by massive data with such high resolution, where the data processing is computational demanding. In this regard, CNN, which is the state-of-the-art method for many big data analysis applications, are exploited as the most important part of our proposed framework. Additionally, CNN is much superior to other approaches in terms of accuracy and efficiency (Hua et al., 2020).

The building detection in our research is actually a semantic segmentation task in the computer vision field. Semantic segmentation is the task to assign each pixel in an image with a class label (Garcia-Garcia et al., 2017). An enhanced feature representation end-to-end can be learned by the CNN model for solving semantic segmentation problems. Recently, the commonly used CNN architecture for urban semantic segmentation tasks is FC-DenseNet (Jégou et al., 2017), which has shown superiority in accuracy (Shi et al., 2020) and has better capability of feature extraction compared with other networks (Shi et al., 2020). In this study, we exploit FC-DenseNet (see Figure 2) for building detection in our proposed framework, which could discriminate "building" and "non-building" for each pixel.

FC-DenseNet extends the DenseNet (Huang et al., 2017) architecture to fully convolutional networks for semantic segmentation. In the DenseNet block, all preceding features are taken as the input, and then its output features are transferred to all subsequent layers. Through the feature reuse, the potential of the network can be yielded that is easier for training and parameter efficiency. Therefore, there are shorter connections within layers close to the input or the output, which enforce the intermediate layers to learn distinguished feature maps. ResNet (He et al., 2016) combines features by summation, which may affect the information flow in the network. Instead, DenseNet combines features by iteratively concatenating them. This contributes to the efficient flow of information in the network for easier training. Moreover, DenseNet layers have small numbers of filters per layer, which means that only a small set of feature maps are added to the network. These attributes of DenseNet allow better parameter efficiency. FC-DenseNet only upsamples the features from the preceding dense block, which have reduced both the amount of computation and the number of parameters. This also makes dense blocks at each resolution of the decoder independent to the number of pooling layers that are used in the encoder. Additionally, the standard skip connection between the encoder and the decoder is used to pass higher resolution information. By reusing features maps, skip connections can facilitate the recovery of spatial details in the decoder from the encoder.

3.2 The Proposed Framework for the Detection of Undocumented Buildings

The undocumented buildings indicate the objects which exist in airborne survey data (nDSM and TrueDOP), but are not recorded

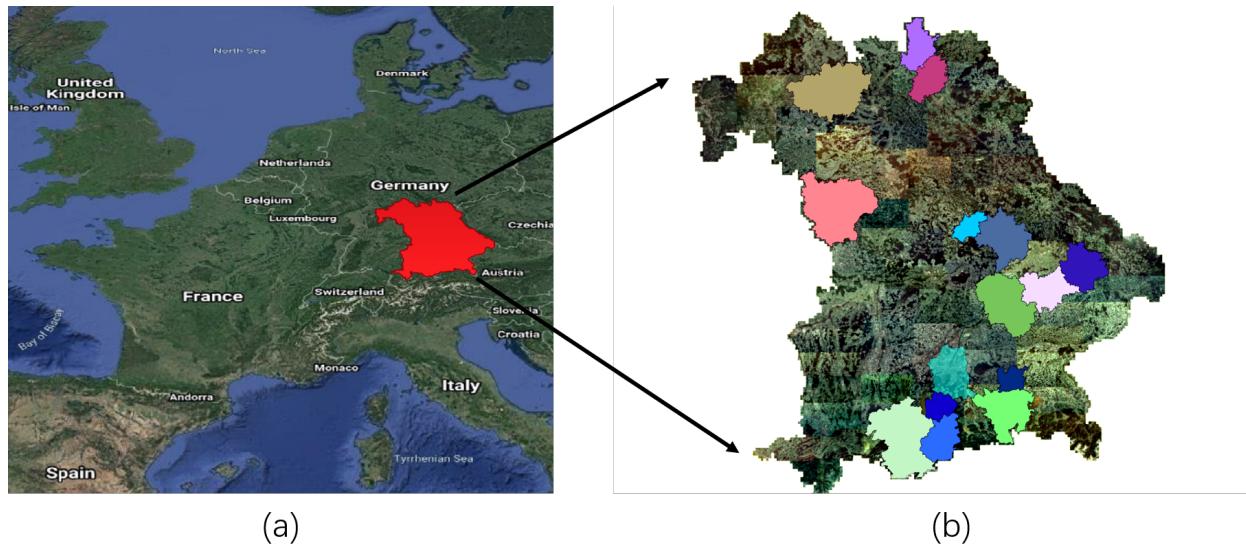


Figure 1. (a)The location of the state of Bavaria, Germany, (b) The study sites in this research which cover 15 districts in the state of Bavaria, Germany.

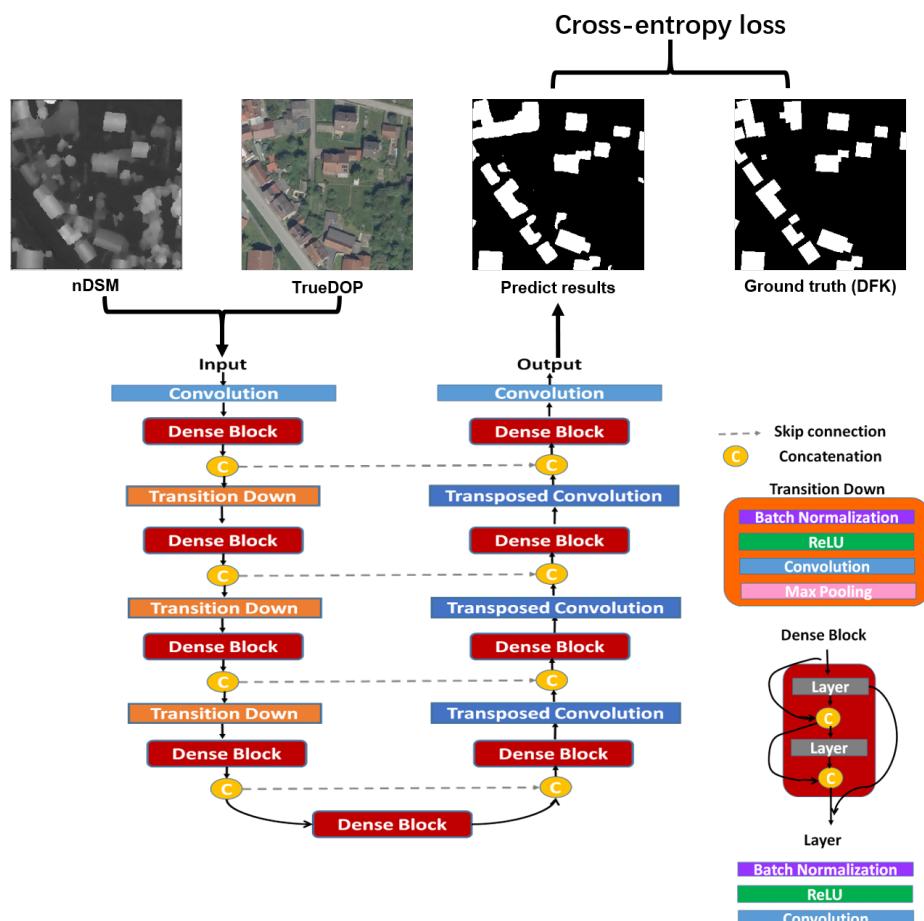


Figure 2. The implemented CNN architectures: FC-DenseNet.

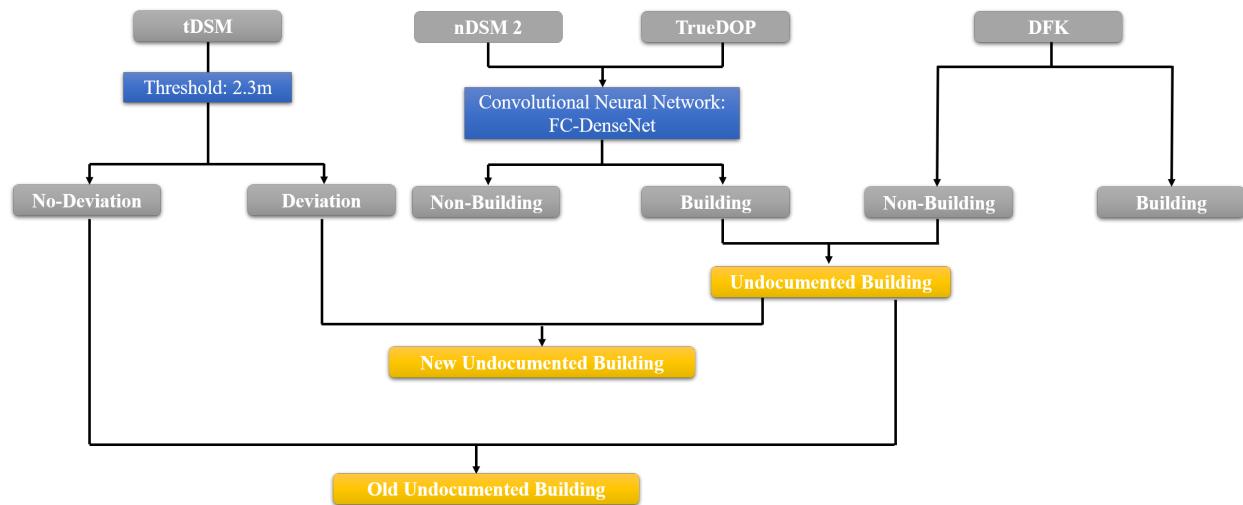


Figure 3. Flowchart of the proposed approach for the detection of undocumented buildings.

in the cadastral two-dimension ground plan (DFK). Utilizing the building detection results, we propose a framework for the detection of undocumented buildings, which is to produce a product of undocumented buildings for 1/4th of the state of Bavaria, Germany using CNN and decision fusion.

An overview of the proposed framework for the detection of undocumented buildings is shown in Figure 3, which can be utilized as a routine strategy for large-scale processing. The proposed framework consists of two main tasks of this study: (1) detection of undocumented buildings, (2) discrimination between old and new undocumented buildings.

The building detection results might be unsatisfactory from individual data source. For instance, automatic building detection from aerial imagery is still limited due to the variation of the appearance of buildings, which results from the atmospheric and seasonal effects, shadow as well as motion blur (Sirmacek, Unsalan, 2008). Buildings and the other impervious objects such as roads share similar spectral and spatial characteristics from aerial imagery. Besides, buildings and other elevated objects above the ground such as trees result in some misclassifications by only using nDSM. Therefore, in our research, we exploit TrueDOP and nDSM together as the input of the CNN model to make a distinction between "building" and "non-building" for each pixel, which can combine the benefits of both radiometric and geometric data. After the predicted results from the CNN model are overlaid with the DFK, the undocumented building pixels can be identified, which relate to pixels those are assigned to the "non-building" in the DFK but are predicted as "building" from the CNN model.

The tDSM is the difference of two DSMs captured at two time points, and with an empiric value (1.8 m), we are able to identify constructional changes if there is a height deviation of this pixel from year 2014 to year 2017. The undocumented building pixels are considered as new undocumented building if there is a height deviation, which indicates that there is a building constructed after time point 1 (year 2014). Otherwise it is regarded as old undocumented building, which means that this building was constructed before year 2014.

4. EXPERIMENTS

4.1 Data Preprocessing

The datasets utilized in this study consist of TrueDOP, nDSM, tDSM, and DFK. Firstly, the tiles of TrueDOP and nDSM which are completely within the district boundary are selected for all 15 districts. For the DFK, they are firstly re-projected to the same projection of the TrueDOP and clipped within the same geo-range. Then, the DFK vectors are rasterized to the raster format. In order to collect enough training patches for large-scale building detection tasks, the TrueDOP and nDSM from 14 districts (exclude Bad Toelz), and the corresponding ground reference DFK are cut into small patches with the size 256×256 pixel, where each patch has an overlap of 124 pixels with its neighboring patches. However, since not all buildings are documented in the DFK, there may be inconsistencies among TrueDOP, nDSM, and DFK when the DFK is utilized as ground reference for the training of CNN. For example, a building is missing in the DFK, while shows in the corresponding TrueDOP and nDSM. This issue is insignificant in our properly selected datasets.

4.2 Experimental Setup

In this study, the CNN model, FC-DenseNet is implemented within a Pytorch framework on an NVIDIA Tesla P100 GPU with 16 GB of memory. The network is trained from scratch using a stochastic gradient descent (SGD) optimizer with a learning rate of 0.000001. The loss function is cross entropy loss and the batch size is 5. For the FC-DenseNet, there are 12 dense blocks with each having 5 convolutional layers.

Our task is aimed at large-scale processing, thus, the selection of training data is investigated firstly. In this regard, two separate models from different sets of training data are trained to examine the transferability issues. The number of training and validation patches are shown in Table 1. Then, we evaluated these two trained models in two different evaluation districts, which are Ansbach and Bad Toelz, respectively. For Ansbach, the evaluation data is the same as the validation data obtained in Ansbach (18077 patches). However, for the district of Bad Toelz, the evaluation data cover the whole area (without training areas).

Table 1. The numbers of training and validation patches of the dataset for two different trained models.

| Trained model | District | Number of training patches | Number of validation patches |
|---------------|----------------|----------------------------|------------------------------|
| 1 | Ansbach | 67965 | 18077 |
| | Ansbach | 67965 | 18077 |
| | Wolfratshausen | 14982 | 3671 |
| | Kulmbach | 24998 | 5679 |
| | Kronach | 19987 | 5112 |
| | Landau | 34964 | 8733 |
| | Deggendorf | 38454 | 9763 |
| | Landshut | 60957 | 15090 |
| | Muenchen | 88364 | 22213 |
| | Regensburg | 47947 | 11941 |
| | Hemau | 9481 | 2243 |
| | Rosenheim | 59141 | 14789 |
| | Wasserburg | 14150 | 3567 |
| | Schweinfurt | 54951 | 13759 |
| | Weilheim | 76959 | 19202 |

5. RESULTS

The building detection results from FC-DenseNet are evaluated using three metrics, which includes overall accuracy, F1 score and intersection over union (IoU). The definitions of these accuracy metrics are described below:

$$\text{Overall accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (5)$$

where TP (true positives) is the number of correctly identified building pixels, FN (false negatives) denotes the number of missed building pixels. FP (false positives) is the number of non-building pixels in the ground reference but is mislabeled as buildings and TN (true negatives) denotes the correctly detected non-building pixels. F1 score is a measure which represents a balance between precision and recall.

5.1 Evaluation Results in Ansbach

Table 2 lists the statistical accuracy of building detection using two different trained models evaluated in Ansbach, and two examples which highlight the performance of the trained model 1 (training data is collected only from Ansbach) are shown in Figure 4. Compared to Figure 4 (b), Figure 4 (a) could also delineate some small buildings.

Several examples of old and new undocumented buildings detected by our propose framework in Ansbach are illustrated in Figure 5 and Figure 6, respectively. Many buildings which are undocumented in the DFK, have been successfully detected by our proposed framework. This has proved the potential of our proposed framework for the detection of undocumented buildings.

5.2 Evaluation Results in Bad Toelz

After reviewing the results evaluated in Ansbach from two trained models with different datasets, we then investigate the

performances of these two trained models in Bad Toelz. It should be noted that Bad Toelz is excluded in neither training data nor validation data for both trained models. This is a more realistic test for the task of large-scale building detection, which could be implemented to upscale the existing training for the building detection results in the whole state of Bavaria, Germany scale. Evaluation in Bad Toelz (see Table 3) shows a great improvement of 12.6% and 17.5% for F1 score and IoU when the training data are collected from 14 districts in the state of Bavaria, Germany rather than only in the district of Ansbach. Figure 7 shows some visual examples for comparison.

Figure 8 and Figure 9 illustrate the results of old and new undocumented buildings in Bad Toelz, respectively. The trained model 2 can differentiate "building" and "non-building" better than the trained model 1. For instance, from the first example in Figure 8, the ground is mislabeled as an old undocumented building (see Figure 8 (a)). However, this false alarm could be avoided by the trained model 2 (see Figure 8 (b)). In addition, the trained model 2 can detect more buildings than the trained model 1, as shown in the first case in Figure 9. In this example, the building which is occluded by the trees could be essentially identified as a new undocumented building by the trained model 2 (see Figure 9 (b)), where the training data covers 14 districts in the state of Bavaria, Germany.

6. DISCUSSION

For Ansbach, the accuracy of building detection is a little higher by just collecting its local training dataset and training a local model. This is due to the fact that in the trained model 1 which is evaluated in Ansbach, the evaluation data share the similar data distribution with training data. Therefore, the trained model 1, which has already a best fit for the training data in Ansbach, achieves better accuracy when the evaluation data is also collected from Ansbach. However, local training is time consuming and expensive if we want to generate the building detection results for the all districts in 1/4th of the state of Bavaria, Germany. Both trained models 1 and 2 achieve comparable results for the detection of undocumented buildings in Ansbach, which implies that if the evaluation data is seen from the trained model, there is nearly no difference when the training data is collected from the local training dataset or a larger dataset.

The results in Bad Toelz have demonstrated the superiority of larger datasets for large-scale building detection tasks when the evaluation district is different from training districts. There are

Table 2. Accuracy of two different trained models evaluated in Ansbach

| Trained model | Overall accuracy | F1 score | IOU |
|---|------------------|----------|--------|
| 1 (training data collected only from Ansbach) | 98.9 % | 90.5 % | 82.7 % |
| 2 (training data collected from 14 districts exclude Bad Toelz) | 98.8 % | 90.3 % | 82.3 % |

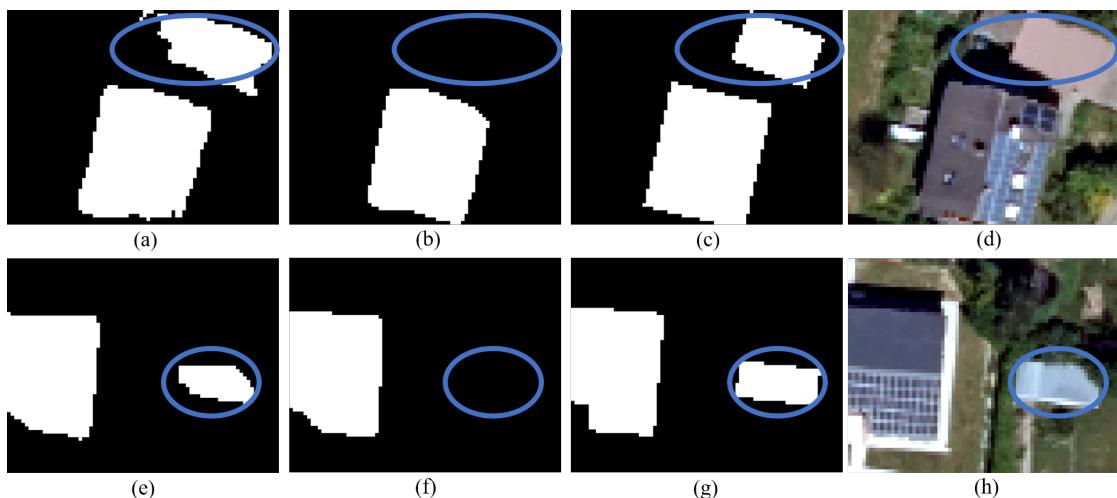


Figure 4. The buildings in Ansbach obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as ground reference as well as (d), (h) TrueDOP.

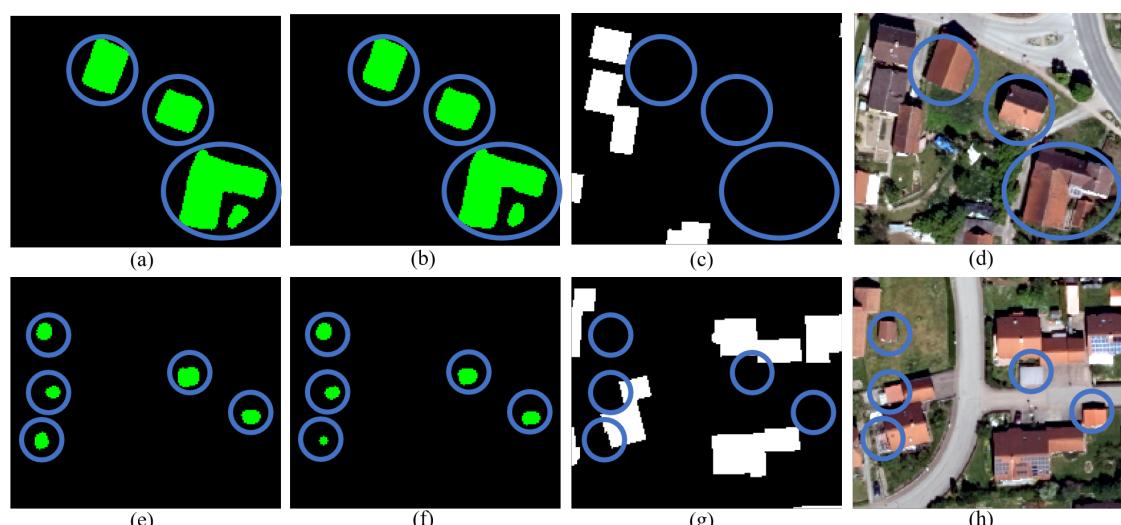


Figure 5. The new undocumented buildings in Ansbach obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as well as (d), (h) TrueDOP.

Table 3. Accuracy of two different trained models evaluated in Bad Toelz

| Trained model | Overall accuracy | F1 score | IOU |
|---|------------------|----------|--------|
| 1 (training data collected only from Ansbach) | 99.6 % | 73.9 % | 58.7 % |
| 2 (training data collected from 14 districts exclude Bad Toelz) | 99.8 % | 86.5 % | 76.2 % |

several reasons for this. One is that the trained model 1, where the training dataset are collected only from Ansbach, might be overfitted also in Ansbach, so it performs not well in Bad Toelz. The other is that the training samples are more representative in the trained model 2, where different types of buildings are included, which is more beneficial to the learning of CNN. In

this regard, using larger dataset could effectively compensate the insufficient information from the limited number of training data.

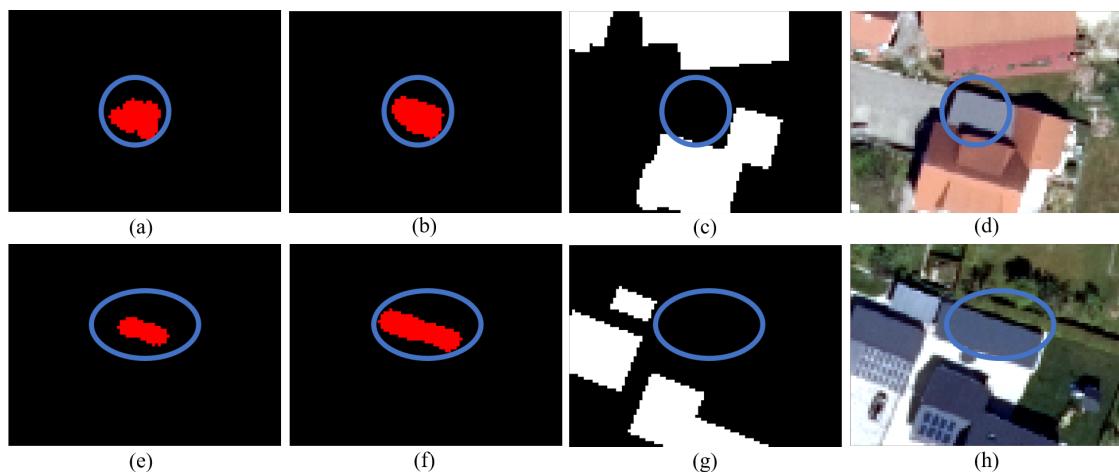


Figure 6. The new undocumented buildings in Ansbach obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as well as (d), (h) TrueDOP.

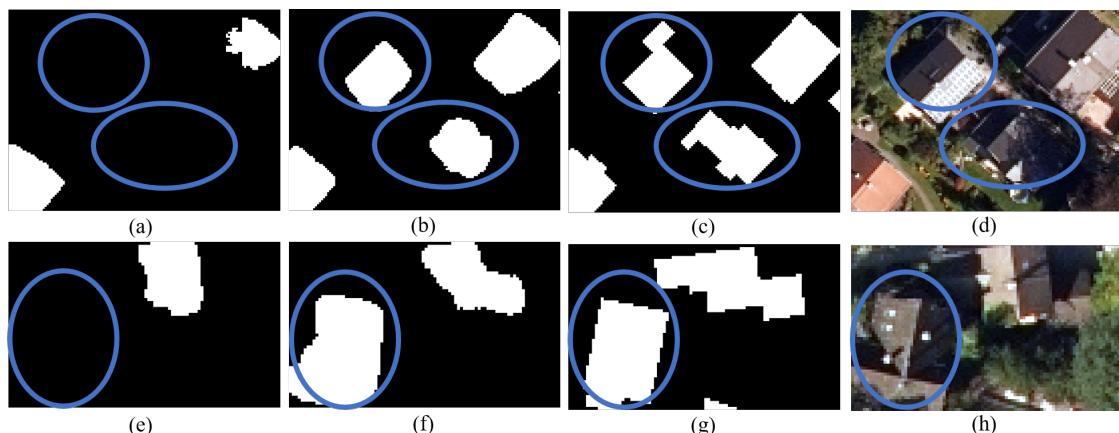


Figure 7. The buildings in Bad Toelz obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as ground reference as well as (d), (h) TrueDOP.

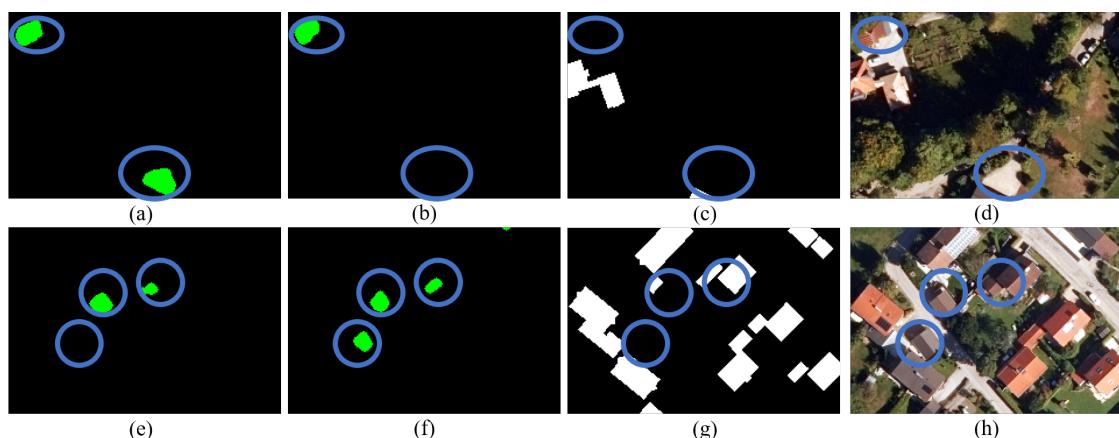


Figure 8. The new undocumented buildings in Bad Toelz obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as well as (d), (h) TrueDOP.

7. CONCLUSION

Considering that the object “building” is one of the most important types of terrestrial objects for real-estate management, we have firstly investigated the transferability issue for the task

of large scale building detection. The model which collects diverse training samples from different districts, has achieved a better accuracy in producing the high resolution (0.4 m) building maps of the state of Bavaria, Germany than that which collects training data only from one district. This impressive result is

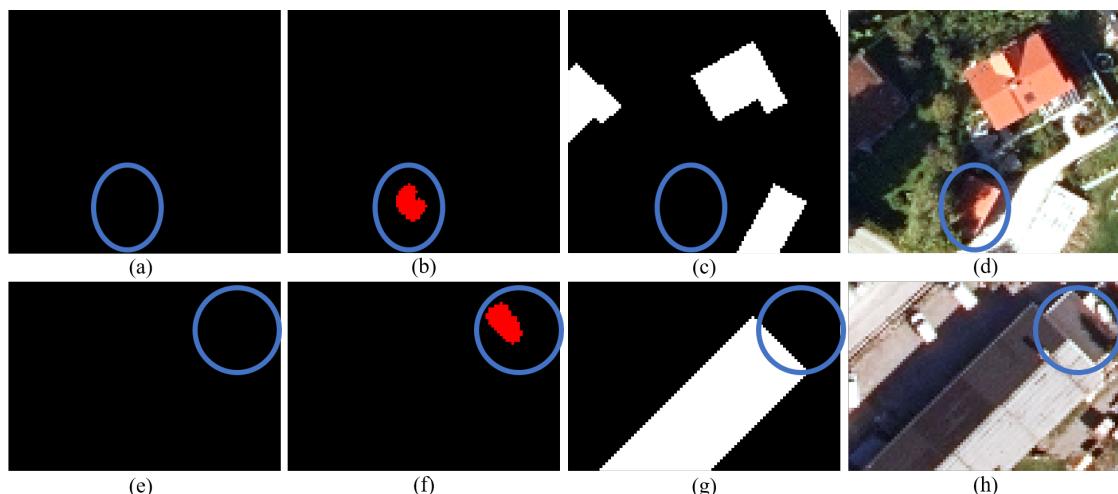


Figure 9. The new undocumented buildings in Bad Toelz obtained from (a), (e) Trained model 1, (b), (f) Trained model 2, and their corresponding (c), (g) DFK as well as (d), (h) TrueDOP.

also beneficial to other large-scale object detection works from remote sensing data. Based on the building detection results, we have proposed a framework for the detection of undocumented buildings from normalized Digital Surface Model (nDSM), orthophoto with Red, Green, Blue bands (TrueDOP), and the corresponding existing Digital Cadastral Map (DFK), which indicates the Convolutional Neural Network (CNN) has the great potential for updating building models in geographic information systems. Moreover, the undocumented buildings can be classified into two types: old or new undocumented building with the aid of a Temporal Digital Surface Model (tDSM). In the future, more possibilities of extension of this work could be investigated, e.g., investigating the performance of introducing the near infrared (NIR) band for building detection.

ACKNOWLEDGEMENTS

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. ERC-2016-StG-714087, acronym: So2Sat, www.so2sat.eu), the Helmholtz Association under the framework of the Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de) and Helmholtz Excellent Professorship "Data Science in Earth Observation - Big Data Fusion for Urban Research". This work is also part of the project "Investigation of building cases using AI" funded by Bavarian State Ministry of Finance and Regional Identity (StMFH) and the Bavarian Agency for Digitization, High-Speed Internet and Surveying.

REFERENCES

- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J., 2017. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.
- Geßler, S., Krey, T., Möst, K., Roschlaub, R., 2019. Mit Datenfusionierung Mehrwerte schaffen – Ein Expertensystem zur Baufallerkundung. *DVV-Mitteilungen*, 2, 159–187.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hua, Y., Mou, L., Zhu, X. X., 2020. Relation Network for Multilabel Aerial Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.
- Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 11–19.
- Li, Q., Qiu, C., Ma, L., Schmitt, M. et al., 2020. Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine. *Remote Sensing*, 12(4), 602.
- Roschlaub, R., Möst, K., Krey, T., n.d. Automated Classification of Building Roofs for the Updating of 3D Building Models Using Heuristic Methods. *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88.
- Shi, Y., Li, Q., Zhu, X. X., 2020. Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 184–197.
- Sirmacek, B., Unsalan, C., 2008. Building detection from aerial images using invariant color features and shadow information. *2008 23rd International Symposium on Computer and Information Sciences*, IEEE, 1–5.
- Vargas-Muñoz, J. E., Lobry, S., Falcão, A. X., Tuia, D., 2019. Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *ISPRS journal of photogrammetry and remote sensing*, 147, 283–293.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.