

AUTOMATED SEMANTIC MODELLING OF BUILDING INTERIORS FROM IMAGES AND DERIVED POINT CLOUDS BASED ON DEEP LEARNING METHODS

E. Gülch^{1,*}, L. Obrock¹

¹ University of Applied Sciences Stuttgart (HFT), Schellingstr. 24, D-70174 Stuttgart, Germany

Commission II, WG II/4

KEY WORDS: Semantic Interior Modelling, Point clouds, Deep Learning, BIM

ABSTRACT:

In this paper, we present an improved approach of enriching photogrammetric point clouds with semantic information extracted from images to enable a later automation of BIM modelling. Based on the DeepLabv3+ architecture, we use Semantic Segmentation of images to extract building components and objects of interiors. During the photogrammetric reconstruction, we project the segmented categories into the point cloud. Any interpolations that occur during this process are corrected automatically and we achieve a mIoU of 51.9 % in the classified point cloud. Based on the semantic information, we align the point cloud, correct the scale and extract further information. Our investigation confirms that utilizing photogrammetry and Deep Learning to generate a semantically enriched point cloud of interiors achieves good results. The combined extraction of geometric and semantic information yields a high potential for automated BIM model reconstruction.

1. INTRODUCTION

The digitalisation of the building sector is progressing steadily and, with Building Information Modeling (BIM), is taking the step from two-dimensional plans on paper to comprehensive, three-dimensional digital building models. These BIM models are the central element and represent the entire life cycle of a building, from planning and operation to demolition. In addition to the three-dimensional component and object geometries, they also contain all relevant semantic information. However, the introduction of BIM is currently taking place virtually only in the planning of new buildings. Due to the very high complexity of manual data acquisition and processing, the recording of already existing buildings as BIM models has been a minor topic of interest so far. Developing an automatic extraction of the necessary information out of measurement data yields a high potential at simplifying the creation of such "As-Built" or "As-Is" models and thus the possibility to make them widely available. Creating these models requires three-dimensional data. We consider photogrammetry as a great method to not only capture and reconstruct buildings as point clouds of high quality but also extract further semantic information out of these images. Especially the categories of objects need to be available, as these are the foundation of every model.

In this paper, we present our approach to providing both semantic and geometric information of an interior room in a classified point cloud in an automated process.

2. RELATED WORK

Building Information Modeling is a major focus of the digital transformation of the building sector. In Germany its implementation is taking place supported by guidelines and investigation mainly focussing on its introduction and execution in various disciplines, e.g. (Egger et al., 2013), (Eschenbruch et al., 2014), (Kaden et al., 2019) and (Bramann et al., 2015b). With the "Stufenplan Digitales Planen und Bauen" the Federal Ministry of Transport and Digital Infrastructure gradually introduced the BIM method into the planning processes of public

infrastructure (Bramann et al., 2015a). The effects of BIM on the implementation of infrastructure projects are being examined in detail during the introductory phase.

A reconstruction of existing buildings as three-dimensional BIM models is possible using measurements of geodetic instruments, e.g. (Borrmann et al., 2015) and (Clemen and Ehrich, 2014). The increased demand of geometric three-dimensional data and additional semantic information of a BIM model is mostly ignored. In consequence, the acquisition and modelling of measurement data is highly complex and requires a large expenditure of time and money.

The whole field of Deep Learning has become a major focus of research in recent years. Especially computer vision based on Convolutional Neural Networks progressed a lot since (Krizhevsky et al., 2012) was able to achieve great improvements in Image Classification. A lot of different ideas further developing this approach, and in turn, improving the reached accuracies, have been published, e.g. (Szegedy et al., 2015), (Simonyan and Zisserman, 2015), (He et al., 2015), (Xie et al., 2017), (Huang et al., 2017). In addition to the pure classification of entire images, the idea of localising the detected class in the image also gained a lot of interest. Two different approaches have emerged for this purpose. Object Detection, as presented in (Girshick et al., 2014), (Girshick, 2015) and (Ren et al., 2015), uses bounding boxes to locate and classify objects. Even more precise, the Semantic Segmentation classifies every pixel of an image. As the resolution of the input images is retained, this method is very well suited to project the extracted semantic information into a point cloud. One of the first popular networks was (Long et al., 2015), which is named Fully Convolutional Network (FCN). Continuing from there, a lot of improved network architectures were developed, e.g. (Jégou et al., 2017) and (Zhao et al., 2017). In this paper we are using DeepLabv3+ (Chen et al., 2018), which achieves excellent results in benchmarks.

In (Obrock and Gülch, 2018), our first approach to automatically generate a semantically enriched point cloud was published. It was based on the application of Deep Learning for segmentation

* Corresponding author

of eight interior building components and objects using an FCN. Subsequently, the segmented images were indexed and inserted into the original image by replacing the blue channel. Based on these false colour images a point cloud was created using photogrammetric methods. The fact that the previously segmented object categories are contained in the colour values of the points, and thus the semantic information confirmed the feasibility of this approach. The steps taken to improve and further develop this approach are presented in this paper.

3. METHODOLOGY AND EVALUATION

3.1 Overview

Images are the main component of our approach because of the massive amount of information contained in them. They are the input for the photogrammetric point cloud generation and the extraction of objects based on Deep Learning methods.

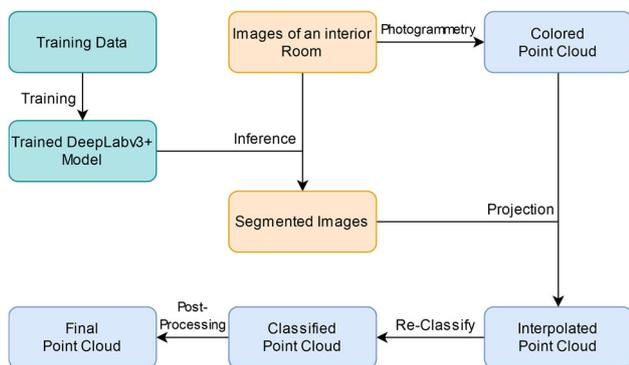


Figure 1. Overview of our workflow. Images of an interior room are the input for inference with our trained neural network and the photogrammetric point cloud generation. An interpolated point cloud is produced by projection of the segmented images. This point cloud is reclassified and used as the input in the final post-processing step.

We use DeepLabv3+ as the architecture of our neural network to segment components and objects of interiors visible in the images at pixel-level. The training of the model is conducted using manually segmented ground truth data.

The trained model then is used for inference on the images of an exemplary room. The base images also are used to generate a point cloud using photogrammetry. Afterwards, the category information stored in the RGB-values of the segmented images is transferred in the point cloud by projection based on the determined camera parameters. Ideally, the result would have been a classified point cloud, but because of interpolation, there is no clear assignment of category colours to the points. Therefore, an additional step is taken to reclassify them. By using the clearly classified point cloud as input for further post-processing, we are able to automatically correct the rotation and scale, as well as extract additional information like floor, ceiling and wall planes.

The presented steps are described in more detail below.

3.2 Semantic Segmentation of Interiors

A comprehensive and high-quality segmentation of all the important building components and objects in the images is of essential importance for a complete reconstruction of an existing building.

For this purpose, the object categories were expanded to a total of 25 components and objects important for a great variety of interiors. Even though “Wall” is a major component of interiors, it had to be dismissed from the segmentation classes of the neural network because it had a negative influence on training. Therefore, the model was trained for extracting the remaining 24 categories and a “Background” class of unclassified objects. As these were to be segmented with Deep Learning out of images, a new training data set was created. It is based on approximately 300 images and the corresponding manually segmented ground truth annotations. Using data augmentation, the training dataset was expanded to almost 18,000 unique images.

Room forming components	Connecting components	Fixed objects of interest		Movable objects of interest
Floor	Door	Light switch	Socket	Poster
Wall	Window	Lamp	Pillar	Bookshelf
Ceiling		Heater	Pipe	Carpet
		Stairs	Railing	Cabinet
		Sink	Toilet	Chair
		Cable trunking	Fire alarm	Table
		Fire alarm siren		Fire extinguisher

Table 1. Overview over the 25 categories of interiors we are interested in. The category “Wall” had to be excluded from the segmentation classes as it couldn’t be segmented by the neural network. Our model was trained for the remaining 24 classes plus a “Background” class of unidentified objects.

To further improve the quality of the segmentation, the architecture underlying the Deep Learning model was exchanged. Instead of a Fully Convolutional Network, the Architecture of DeepLabv3+ was used as the basis for training. Based on a pre-trained model utilizing xception65 architecture, training was conducted using fine-tuning.

In an additional post-processing step, the smallest areas, which are partially present in the segmentations but do not relate to any real object, are filtered out.

Based on the final trained neural network, inferencing was conducted on images of an interior room as the basis for subsequent steps.

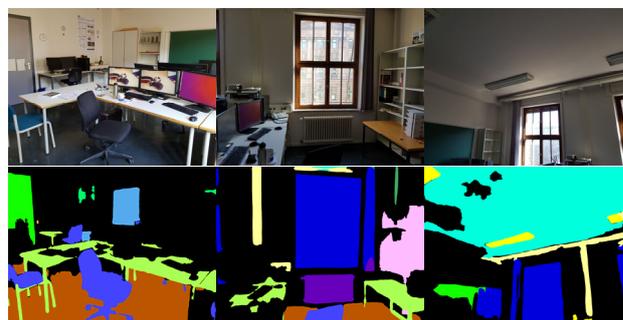


Figure 2. Figure 2: Exemplary results of using our trained DeepLabv3+ model for inferencing the images of an interior room.

The segmentations achieved seem to be matching very good despite the significantly increased complexity due to the expansion of the categories.

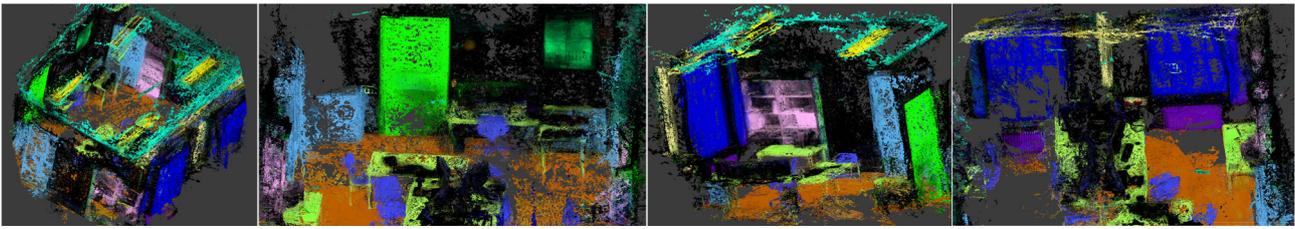


Figure 3. Different views of the generated point cloud after projection of the categories as combination of colour values. No clear distinction of categories, as the colours changed through interpolation, which is especially apparent in parts of the green door.

Due to time constraints, we were unable to create another dataset for validation and thus could not calculate the accuracy of the model.

Visually inspecting the images, the overall quality seems to be rather high. Some of the categories, especially ones with large areas, seem to have a high consensus. Errors are happening where objects resemble each other too closely, like the feet of the “Table” and “Chair”. It is also noticeable, that unique objects like “Heater” are segmented more precisely than plain and big areas like “Ceiling”, in which some holes occur. The small categories like “Light switch” and “Socket” are segmented very unsteady, often depending on the angle and distance at which they are pictured. A continuously only mediocre segmented category is the “Bookshelf”, as its areas seem to be incomplete often. Nevertheless, the results obtained are very promising and form a good basis for the next steps.

3.3 Classified Point Cloud

With BIM-models as our main target, we then need to transfer the two-dimensional semantic information from the images into three dimensions. Using photogrammetry and digital image matching, the images of the room build the basis to create a three-dimensional, semantically enriched point cloud.

In our previous paper, the category information was compressed into just one colour channel, which made it hard to distinguish between them visually because there was only a small distance between the classes. This would have been further complicated by increasing the classes to 25, including the unclassified “Background”. Therefore, we needed to change the approach. The transfer of semantic and geometric information no longer takes place in one, but in two separate but consecutive phases. Based on these, a photogrammetric point cloud is generated automatically in Agisoft Metashape (Agisoft, 2020) without placing any control points. In our experiments this software has proven to be very well suited to perform this task, especially in comparison to some non-commercial packages, like Visual SfM or Alice Vision.

By relying on the original images of the interior room, a point cloud is generated without losing any colour information. To transfer the category information of the segmented images into the point cloud, a different method is applied. By replacing the original images with the segmented images after the point cloud was created and thus relying on the previously determined camera position and rotation, we are projecting the category information into the point cloud. The resulting interpolated point cloud is shown in Figure 3.

This yields significant advantages. When generating the point cloud by photogrammetry, there is no missing colour information. This results in a higher quality when extracting and linking individual concise image areas and thus an improved reconstruction of the entire point cloud. Furthermore, the colour value combinations representing the individual categories can be placed at a greater distance from each other. These are very small if 25 classes are to be divided into 256 possible colour values of one channel. In contrast to this, the combination of values from

three channels allows significantly larger Euclidean distances between the colour values of the categories, since they can be regarded as points in three-dimensional space. A distribution over all colour channels also has the greatest advantage for the viewer, since they can be distinguished more clearly from each other.

The point cloud itself is a proof of concept of an automatic generation and category projection. It is rather noisy but still manages to capture most parts of the room. Nevertheless, there are some missing areas, especially at the floor and the ceiling, which consist of large, uniform parts.

The colour values of the individual points are derived by interpolation of the overlapping values of the individual images. If their camera orientations in three-dimensional space or if the segmentations performed in these images do not match, the interpolation results in divergent values in the generated points. This becomes very clear by the darkening colours within the points of an object like the door, which only is partially bright green.

This leads to the fact that a further step is necessary, in which a reclassification of the point cloud is carried out to generate a clear assignment of the categories.

To aid this reclassification, the RGB colour values of the categories were chosen so, that their minimum Euclidean distance from each other and possible interpolated colours with “Background” are maximally large. This way a reasonably even distribution of the colour values in three-dimensional space is achieved and a clear assignment of the categories is made possible more easily.

A coarse filtering of the point cloud is carried out beforehand to remove outlier points that are not close enough to the other points. A clear assignment of categories is obtained by looking at the colours of the individual points in three-dimensional space and determining their Euclidean distance from the colour values of the individual categories. Additionally, the distances of the colour values of a point to those colour values are calculated, which would result from the interpolation of a category with “Background” points as the most common class. If the distances to one of these are smaller than a threshold value derived as a percentage of the minimum distance between the classes, a direct assignment to the corresponding category is made.

In the case that these two are not within the limits, the neighbourhood of the points is examined for the assignment of the categories. If one category occurs significantly more often than the others in this neighbourhood, this category is also used for the examined point. As a combination, it is checked if the colour values of a point are within an extended range to the most prominent category in the neighbourhood. If no categories can be determined by distance and neighbourhood investigations, they are listed as an undefined “background” point. The point cloud resulting from this step now contains clearly assigned categories for each point. The classified point cloud is shown in Figure 4.

It becomes clear that due to the overlapping of many segmented images, individual segmentation errors in these images rarely influence the final classification of the point cloud. Incorrect category assignments usually occur if there are systematic errors

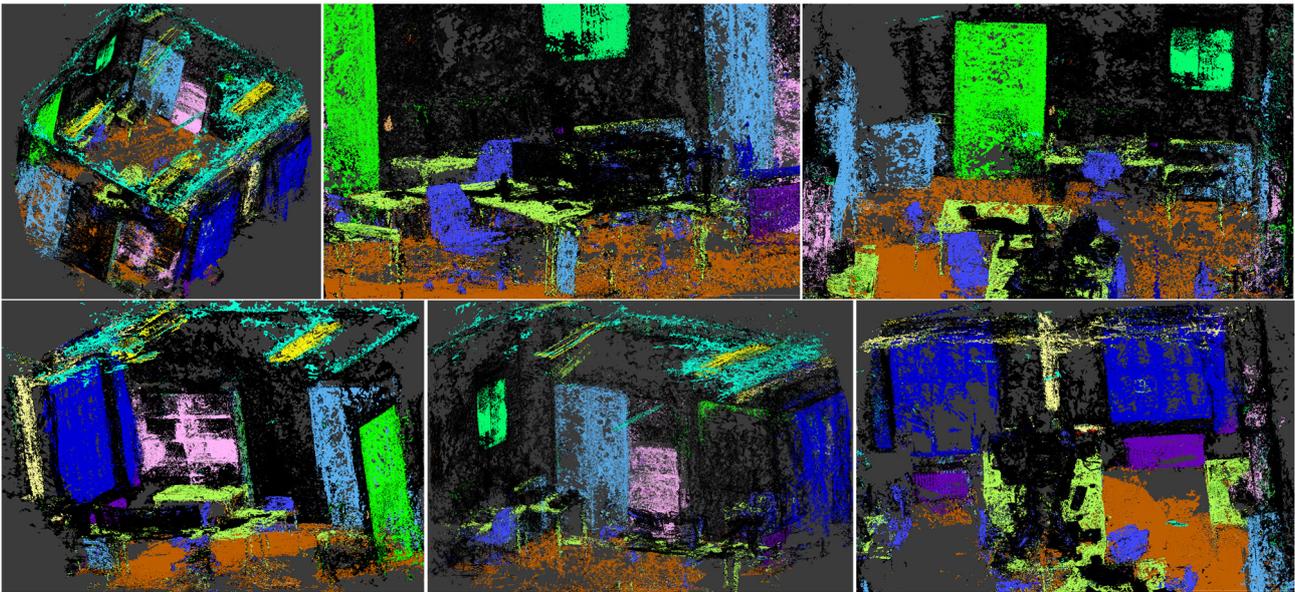


Figure 4. Different views of the classified point cloud. A sharp distinction between categories is visible. Therefore, there are no more darkened colours in the area of the green door.

in the segmented images, the determined camera position and rotation are inaccurate, or image matching errors occur during point cloud creation.

In Table 2 the accuracies are shown, which reflect the good overall classification quality. It is based on the comparison with a manually segmented copy of the point cloud and enable us to extract data about the accuracy (ratio of correctly classified points to all / category points) and mean intersection over union (mIoU) of the categories present in the room.

The good quality of the classification is confirmed by the mIoU with a value of 51.9 % where every class is weighted the same. This also applies for the calculated mean accuracy of 60.4 %, which is showing a high consensus of the ground truth points.

Comparing the individual IoU numbers, a big difference between them becomes obvious, ranging from low values of “Light switch” with 6.3 % to high values of “Door” at 77.9 %. Especially bigger objects seem to be classified very good and even the “Bookshelf”, which often is segmented poorly in the images, still reaches an IoU of 38.4 %. On the other hand, the categories of “Socket” with 10.8 % and “Light switch” with 6.3 %, which represent very small objects, only achieve a low IoU in the point cloud. Partly contrary to this, the accuracy of “Light switch” is reaching a much better value with 48.2 %. This is due to a comparably high amount of interpolated and then wrongly assigned colours which are occurring at points of the floor. Because the number of actual points of “Light switch” is small, these wrongly assigned points have a huge influence on the IoU. In conclusion, it can be said that thought projection and reclassification of the categories in the point cloud a high quality of classification is reached.

3.4 Automated Post-Processing

Based on the resulting classified point cloud, further processing is conducted. Since it was generated fully automatically and without the use of control points and marks, neither its rotation in space nor its scale correspond to the real conditions. Extracting these is possible based on the additional semantic information.

For this purpose, the points segmented as ground are used and occurring incorrectly classified points are filtered out. From the remaining selected points, the ground plane is derived. Then the point cloud is shifted to the origin and rotated in an iterative process based on the ground plane so that its normal is approximately identical to the Z-axis of the local coordinate system and therefore aligned horizontally. Next, the ceiling plane is determined from the corresponding points segmented as ceiling.

Subsequently, wall planes and wall points are extracted from the point cloud as they could not be considered in the segmentation. This is achieved based on a top view heat map created from the unclassified “Background” points of the point cloud, where the pixel values are based on the numbers of existing points of a grid. When viewed in two dimensions, an accumulation of points is to be expected, especially on vertical components. Since walls, as the limiting element of the room, have large vertical surfaces, a particularly strong accumulation is to be expected.

	Mean	Background	Floor	Door	Window	Light switch	Socket	Lamp	Heater	Poster	Cable trunking	Cabinet	Ceiling	Bookshelf	Table	Chair
IoU	51.9	60.8	75.1	77.9	75.6	6.3	10.8	59.2	66.4	43.6	41.6	71.4	53.5	38.4	43.3	54.9
Accuracy	60.4	93.8	77.1	84.4	88.8	48.2	11.3	62.2	70.5	45.3	43.0	74.2	54.7	38.8	50.0	63.4

Table 2. Calculated intersection over union and accuracy of the objects present in the room.

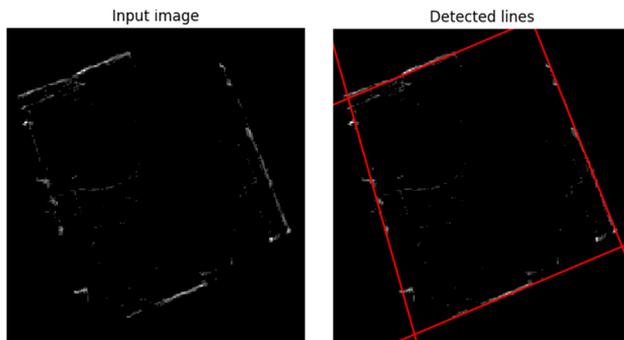


Figure 5. Derived heat map of the room used to extract the most prominent lines using Hough transform algorithm.

From the created heat map, the probable angles of the walls are extracted based on the most prominent lines determined by a Hough transform algorithm. These are then transferred back into the third dimension. Previously uncategorized points, which are located in only a short distance from these, are now selected to determine the actual best-fit planes of the walls. Accordingly, the points close to these planes are classified as wall points.

To achieve a correct scale for the point cloud, the dimensions of an object in the real world and a segmented object in the point cloud have to be adjusted. As an object of comparison, especially doors (frames) seem to be well suited. They are present in every room and easily measurable manually in the real world. They have a rather large dimension in Z-axis direction but are usually easier to measure, than e.g. the vertical distance between floor and ceiling, whose planes in the point cloud are known but may not be completely parallel, making the height calculation error prone. Doors, in contrast, are clearly segmented in the point cloud. Because the point cloud is correctly aligned, the height of objects can easily be extracted and is therefore used for comparison. Using a region growing approach all points of a door are selected and the height is calculated based on the maximum and minimum values along the Z-axis. The resulting scale is used to adjust the point cloud.

When using the scaled point cloud for comparison, the measured distances are matching the real distances rather well as shown on two examples in Table 3.

	Real World	Point Cloud
Height Table	72.3 cm	72.1 cm
Width Cabinet	110.0 cm	108.9 cm

Table 3. Comparison of measurements of objects in the real world to their counterparts in the scaled point cloud.

The points resulting from the intersections of walls with floor and ceiling planes, all of them automatically derived entirely from data, represent the room geometry in its basic features.

With our investigations and the solutions based on them, we have taken further important steps to enable automated BIM-compliant modelling of existing buildings.

4. CONCLUSION

In this paper, we were able to verify that the combination of photogrammetry and Deep Learning is a solid approach to generate a semantically enriched point cloud of interiors. The combined extraction of geometric and semantic information based on segmentation with DeepLabv3+ and projection into the photogrammetric point cloud achieves good results. In consequence, the components and objects can be differentiated very well in the point cloud. The reached mIoU of 51.9 % for the classified point cloud confirms the good quality of this approach. Additional important information essential for a BIM model can be extracted by analysing and post-processing the point cloud. We are confident that we will be able to improve the results by further optimizing the methods used.

In the future, it is planned to extend this approach to the combination and joint processing of several rooms as well as including mobile laser scanners. With the methods presented by us, we have created a solid basis for the acquisition and modelling of semantic and geometric information of interiors for BIM models towards their automated reconstruction.

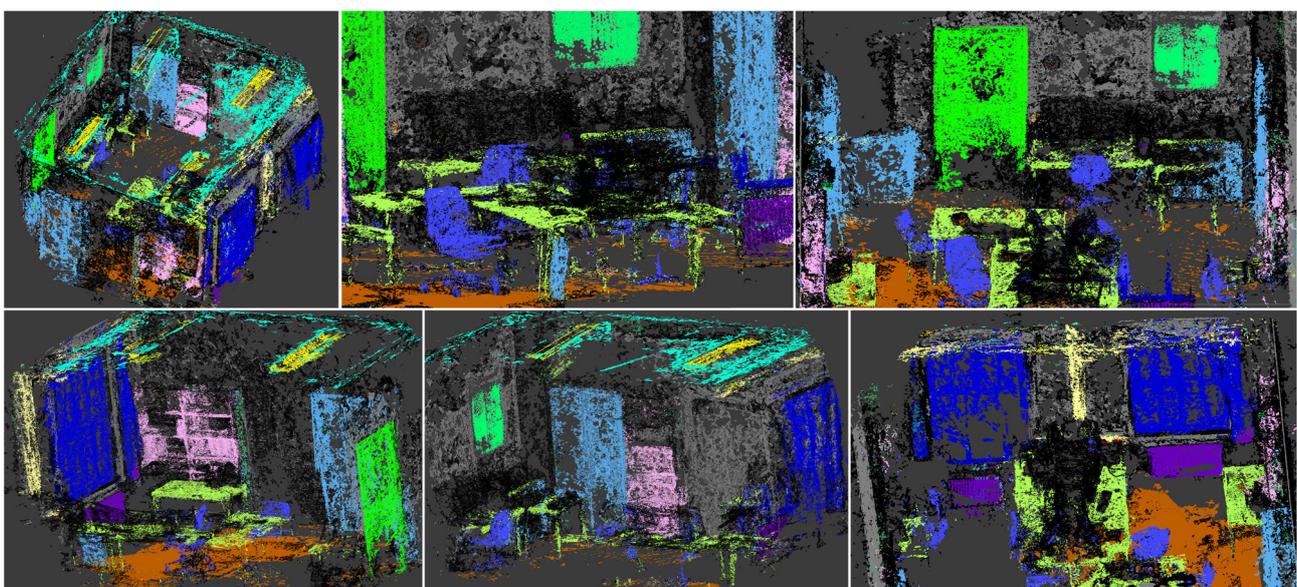


Figure 6. Different views of the post-processed final point cloud. Rotation and scale are corrected. Extracted wall point are visible in the point cloud in light grey colour.

ACKNOWLEDGEMENTS

The project “i_city: BIM-konforme Gebäudeerfassung” is sponsored by the Federal Ministry of Education and Research (BMBF) under the promotion code 13FH9E011A and supervised by the project executing organization VDI Technologiezentrum GmbH for the BMBF. The responsibility for the content lies with the authors.

REFERENCES

- Agisoft, 2020. Agisoft Metashape. <http://www.agisoft.com/> (31 March 2020).
- Borrmann, A., König, M. Koch, C., Beetz, J., 2015. *Building Information Modeling - Technologische Grundlagen und industrielle Praxis*. VDI-Buch, Springer Vieweg, Wiesbaden, 343-361.
- Bramann, H., May, I. and planen-bauen 4.0 – Gesellschaft zur Digitalisierung des Planens, Bauens und Betriebens mbH, 2015a. *Stufenplan Digitales Planen und Bauen*. https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/stufenplan-digitales-bauen.pdf?__blob=publicationFile (31 March 2020).
- Bramann, H., May, I. and planen-bauen 4.0 – Gesellschaft zur Digitalisierung des Planens, Bauens und Betriebens mbH, 2015b. *Konzept zur schrittweisen Einführung moderner, IT-gestützter Prozesse und Technologien bei Planung, Bau und Betrieb von Bauwerken – Stufenplan zur Einführung von BIM*. https://www.bmvi.de/SharedDocs/DE/Anlage/DG/Digitales/bim-stufenplan-endbericht.pdf?__blob=publicationFile (31 March 2020).
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European conference on computer vision (ECCV)*, 801-818.
- Clemen, C., Ehrlich, R., 2014. Geodesy goes BIM. In: *allgemeine vermessungs-nachrichten*. (avn) 121 (6). <https://gispoint.de/artikelarchiv/avn/2014/avn-ausgabe-62014/2552-geodesy-goes-bim.html> (31 March 2020).
- Egger, M., Hausknecht, K., Liebich, T., Przybylo, J., 2013. *BIM-Leitfaden für Deutschland*. https://www.bmvi.de/SharedDocs/DE/Anlage/DG/Digitales/bim-leitfaden-deu.pdf?__blob=publicationFile (31 March 2020).
- Eschenbruch, K., Malkwitz, A., Grüner, J., Poloczek, A., Karl, C. K., 2014. *Maßnahmenkatalog zur Nutzung von BIM in der öffentlichen Bauverwaltung unter Berücksichtigung der rechtlichen und ordnungspolitischen Rahmenbedingungen – Gutachten zur BIM-Umsetzung*. https://www.bmvi.de/SharedDocs/DE/Anlage/DG/Digitales/bim-massnahmenkatalog.pdf?__blob=publicationFile (31 March 2020).
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 580-587.
- Girshick, R., 2015. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440-1448.
- He, K., Zhang, X., Ren, S., Su, J., 2015. Deep Residual Learning for Image Recognition. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q., 2017. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261-2269.
- Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1175 - 1183.
- Kaden, R., Clemen, C., Seuß, R., Blankenbach, J., Becker, R., Eichhorn, A., Donaubaue, A., Kolbe, T. H., Guber, U., DVW – Gesellschaft für Geodäsie, Geoinformation und Landmanagement e. V., Runder Tisch GIS e.V., 2019. *Leitfaden Geodäsie und BIM Version 2.0* (2019). https://rundertischgis.de/images/2_publikationen/leitfaeden/GeoundBIM/Leitfaden%20Geod%C3%A4sie%20und%20BIM_Onlinversion.pdf (31 March 2020).
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems (NIPS 2012)*, 1097 - 1105.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431-3440.
- Obrock, L. S., Gülch, E., 2018. First Steps to Automated Interior Reconstruction from Semantically Enriched Point Clouds and Imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2, 781-787.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems (NIPS 2015)*, 91-99.
- Simonyan, K., Zisserman, A., 2015 Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations (ICLR2015)*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9.
- Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., 2017. Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1492-1500.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2881-2890.