3D RECONSTRUCTION OF SIMPLE BUILDINGS FROM POINT CLOUDS USING NEURAL NETWORKS WITH CONTINUOUS CONVOLUTIONS (CONVPOINT)

M. Kada[1,*]

[1] Institute of Geodesy and Geoinformation Science, Technische Universität Berlin, Germany – martin.kada@tu-berlin.de

Commission IV, WG IV/9

KEY WORDS: Buildings, City Modelling, Point Clouds, 3D, Reconstruction, Deep Learning, Neural Networks.

ABSTRACT:

The automatic reconstruction of 3D building models from airborne laser scanning point clouds or aerial imagery data in a model-driven fashion most often consists of a recognition of standardized building primitives with typically rectangular footprints and parameterized roof shapes based on a pre-defined collection, and a parameter estimation so that the selected primitives best fit the input data. For more complex buildings that consist of multiple parts, several such primitives need to be combined. This paper focuses on the reconstruction of such simple buildings, and explores the potential of Deep Learning by presenting a neural network architecture that takes a 3D point cloud of a single building as input and outputs the geometric information needed to construct a 3D building model in half-space representation with up to four roof faces like saddleback, hip, and pyramid roof. The proposed neural network architecture consists of a roof face segmentation module implemented with continuous convolutions as used in ConvPoint, which performs feature extraction directly from a set of 3D points, and four PointNet modules that predict from sampled subsets of the feature-enriched points the presence of four roof faces and their slopes. Trained with the RoofN3D dataset, which contains roof point segmentations and geometric information for 3D reconstruction purposes for about 118,000 simple buildings, the neural network achieves a performance of about 80% intersection over union (IoU) for roof face segmentation, 1.8° mean absolute error (MAE) for roof slope angles, and 95% overall accuracy (OA) for predicting the presence of faces.

1. INTRODUCTION

For well over two decades, more and more 3D geodata have been captured on larger scales in addition to conventional 2D geodata. And the topographic objects with a relevant vertical extent (like buildings, bridges, trees, etc.) have been modelled accordingly, both strictly in their geometric form, but also with their semantic structure. 3D city models, which usually consist of terrain, buildings, vegetation, and possibly street furniture, are just one example for 3D geodata. Applications for 3D city models are given, e.g., by Biljecki et al. (2015).

For the automatic reconstruction of 3D building models from airborne laser scanning 3D point clouds or aerial and satellite images, mainly methods using deterministic algorithms have been applied so far. In many areas of geodata interpretation, one observes nowadays the emergence of data-driven and artificial intelligence methods that make use of machine learning and, more recently, Deep Learning using neural networks. However, besides the more general tasks like object classification, object detection, and semantic segmentation, etc. one can hardly find research on the geometric reconstruction of 3D building models from point clouds making use of these recent developments.

In this paper, an attempt is therefore made to fill this gap and derive and model the 3D geometry of simple shaped buildings based on airborne laser scanning 3D point clouds using neural networks. Simple shaped buildings are considered to be those that feature a roughly rectangular footprint and a standard roof shape (saddleback, hip, pyramid, etc.) without any roof superstructures. These buildings are often reconstructed using primitives with parameterized roof shapes, by identifying the roof type and estimating the roof parameters that best fit the input data. Here, a neural network architecture is proposed that performs 3D building reconstruction by predicting a semantic segmentation of the input point cloud that labels the roof points according to the roof faces they are located in, also referred to as part segmentation, and by predicting the roof geometry in form of up to four roof face slopes. Based on the (rectangular) footprint, the geometries of 3D building models are constructed using half-space modelling, where a (convex) 3D polyhedron is defined by a set of intersecting planes.

Besides a 3D reconstruction of simple buildings, the proposed approach can be considered the first step towards a model- or primitive-based reconstruction of complex building shapes that are generated from several parameterized primitives, as it has been pursued in 3D building reconstruction for many years. For this purpose, the proposed neural network module for simple buildings could be embedded in an extended architecture for object detection that identifies and localizes the building parts, and forwards the detected oriented bounding boxes to the proposed reconstruction module to derive the geometries of the therein contained building primitives. Object detection neural networks for 3D point clouds, like VoteNet (Qi et al., 2019), are potential foundations for such a comprehensive 3D building reconstruction pipeline.

2. RELATED WORK

The topic of 3D reconstruction of buildings from airborne laser scanning or aerial or satellite images has interested researchers for more than two decades. Haala and Kada (2010), e.g., give a thorough overview on this topic. The proposed approaches are often differentiated into model-driven and data-driven (Maas

* Corresponding author

and Vosselman, 1999). In model-driven approaches, parametric roof templates are typically assumed and the methods estimate the parameter values and roof types that best fit the input data. For buildings with rectangular footprints, Henn et al. (2013), e.g., apply a robust estimation of roof parameters to generate 3D building model hypothesis for ten standard roof types, and verify the best one by a supervised learning approach. Nguatem et al. (2013) use importance sampling to derive the roof shape parameters. Since buildings often have more complex shapes, the respective building models need be composed of several such primitives. Haala et al. (1998) decompose the given footprints into overlapping rectangles, and determine for each rectangle the best fitting parametric roof shape primitive. Kada and McKinley (2009) construct cell decompositions of building footprints for this purpose. Based on a roof face segmentation, Verma et al. (2006) as well as Oude Elberink and Vosselman (2009) construct for each building a roof topology graph, and identify the roof shapes by subgraph matching. In data-driven 3D building reconstruction approaches, fewer assumptions are made about the building (sub-)shapes, and the 3D models are constructed and assembled from a set of low-level geometric primitives, typically planes, that are detected in a previous step. From a set of candidate faces that result from a pairwise intersection of extracted planes, Nan and Wonka (2017) use a binary linear programming approach to find a suitable subset of faces that describes a manifold and watertight polyhedral 3D building model.

In this paper, the focus is on the 3D reconstruction of simple buildings with rectangular footprints for which the roof shape parameters are determined by predicting the slope and position of up to four planar roof faces from a 3D point cloud. For this purpose, a Deep Learning approach is proposed using point cloud based neural networks. For more complex buildings, the proposed reconstruction module can be combined with a footprint decomposition approach or integrated into an object detection neural network that finds and localizes the necessary building components. Although Deep Learning is a large field, the interest for the proposed method is on neural networks that work directly on 3D point cloud data, and do not need to transform the input data into another representation first. In the past years, several neural network architectures have been proposed for object classification, part and semantic segmentation like PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), PointConv (Wu et al., 2019), PointCNN (Li et al., 2018), KPConv (Thomas et al., 2019), and ConvPoint (Boulch, 2020). Besides the mentioned tasks, these networks can be used for point-wise feature extraction in an extended architecture. In the proposed approach, ConvPoint is used, as it defines convolutional filters for 3D point clouds, which have proven to be effective for feature extraction, and allows at the same time comparatively elegant network architectures.

In recent years, more and more approaches based on neural networks have been proposed that reconstruct the geometry of 3D building models. Mahmud et al. (2020) extract building footprints and pixel-wise heightmaps from single overhead images, and produce 3D block models of buildings with a median height. Alidoost et al. (2019) extract building roof lines and heights from images to construct block models as well as simple parametric models with standard roof shapes. Qian et al. (2021) propose a generative adversarial network (GAN) called Roof-GAN to construct the geometry of residential roof structures that are always aligned with the coordinate axes, which might be considered a limitation. A more thorough overview on Deep Learning approaches for 3D building reconstruction is given in (Buyukdemircioglu et al., 2021).

3. NEURAL NETWORK ARCHITECTURE

The neural network architecture proposed in this paper takes as input a cutout of an airborne laser scanning 3D point cloud consisting mainly of the roof points of a single building, and produces a pointwise segmentation of the roof faces as well as the necessary information to construct a 3D building model in a half-space representation. Following the two main objectives, the architecture consists of two main components, outlined in Figure 1, and explained in more detail below. As mentioned in the introduction, it is assumed that the buildings are simple, i.e. they feature a rectangular footprint, and a roof shape with a saddleback, hip, or pyramid roof.

3.1 Roof Face Segmentation

The first component of the neural network takes as input a 3D point cloud that can contain any number of points and assigns a label to each point. The objective thereby is that all points that belong to the same planar roof face get the same label, which allows to recognize the constituent parts of the building roof. This task is commonly considered a semantic segmentation at shape level and is also referred to as part segmentation. Although the segmentation of roof faces is not intended to be a primary output of the neural network, and should in the future also not be directly used in the construction of the 3D building model, itself, it is still required as an intermediary byproduct within the neural network, and is also critical to successfully train the network.

Both semantic and part segmentation are essentially point-wise classification tasks. Since the three supported roof shapes have at most four faces, it is sufficient to differentiate between five classes, where four classes are used for the roof faces and one for all other points that do not actually belong to the roof. Since the network must somehow distinguish the roof points based on some geometry derived features, the four roof point classes are defined according to the orientation of the faces that the points lie within. Here, the face orientation is considered to be the horizontal component of the normal vector of the plane spanned by the face. If the normal vector is roughly directed to the northwest, northeast, southeast, and southwest, then the points of the corresponding faces are labeled with the class labels 1, 2, 3, and 4, respectively, or with the class label 0 otherwise. Thus, the 360° of possible face orientations are divided into four quadrants, so that the class labels can be quickly assigned based on the signs of the x and y components of the normal vectors; although any other definition of four quadrants is also feasible. Obviously, difficulties in the correct prediction of roof classes are to be expected for those roof faces with an orientation exactly on or near the boundaries between two quadrants.

For the roof face segmentation task, a simplified version of the part segmentation neural network of ConvPoint is used that defines continuous convolutional layers for 3D point cloud data (Boulch, 2020). These continuous convolutional layers perform, similar to the convolutional layers in convolutional neural networks (CNNs) for discrete 2D image data, a point-wise multi-scale feature extraction from increasing point neighborhoods when stacked in a multilayer architecture. Because the input point clouds being used in the experiments contain comparatively few points with at most 350 points, only three continuous point convolutional layers are used for both sampling and upsampling in a typical U-Net like architecture with skip connections for intermediate feature information. The (down)sampling layers of the encoder thereby pass 64, 16, and 8 points to their next layers, and the upsampling layers of the

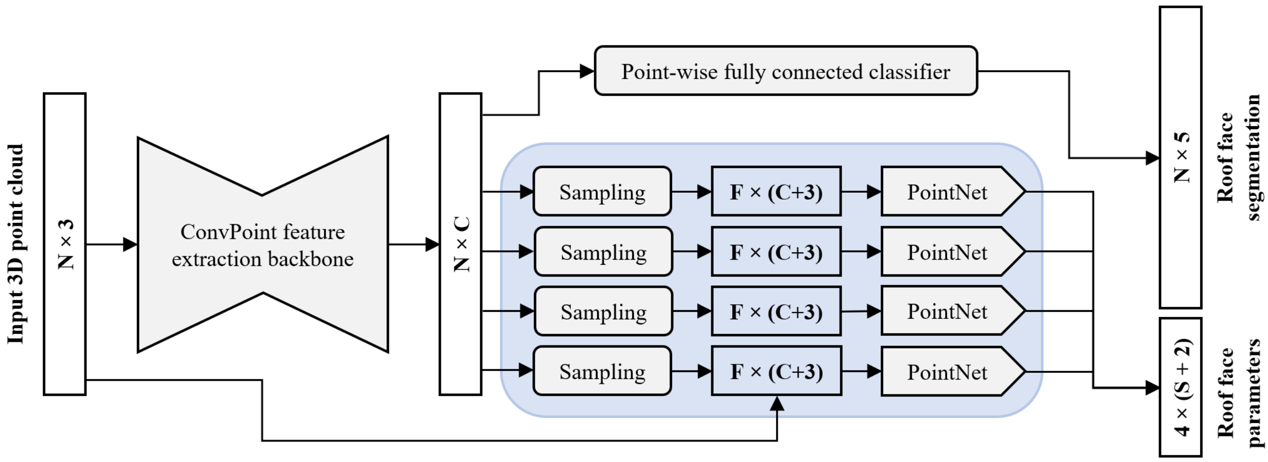


Figure 1. Proposed neural network architecture for the 3D reconstruction of simple buildings from 3D point clouds (N input points, C-dim features, F sampled points per face, S slope angle classes).

decoder similarly propagate the extracted features back to the original point cloud, which results in a feature vector for each input point. All layers are defined with 96 convolutional filters, therefore outputting 96 feature channels. Finally, the last layer is a point-wise linear layer that generates the five class scores for each point. For more details on the part and semantic segmentation network of ConvPoint, see (Boulch, 2020).

3.2 Roof Face Parameters

In the proposed approach, the geometric construction of the 3D building models is performed by using half-space modelling through the definition of planar half-spaces. This requires up to four plane equations per building for the respective roof faces. Since it is assumed that for each building the rectangular footprint is given by exactly four line segments, each of these planes can be determined by the direction of one line segment, a slope (or rotation) angle, and a 3D point through which the plane passes. Since the 3D point is easily determined using the roof face segmentation, e.g. as the mean location of all points that make up a roof face, the only missing information is the slope angle, which is to be determined by the neural network. Because it is generally difficult to precisely predict angles by regression, a similar approach as in (Qi et al., 2019) is taken, where bounding box angles are predicted as a combination of discrete angle classes and continuous correction values.

The four slope angles of the roof faces are determined based on the point-wise features extracted by the U-Net module of the roof face segmentation. But not from all points of the input point cloud, but rather from four subsets of points, for which the roof face segmentation component predicted the points to belong to the respective faces. The slope of the northwest oriented face is, e.g., determined only from those points that are classified to belong to the northwest face. For each of the four face orientations, a total of 32 random points is sampled with replacement, i.e. points can be included several times in a subset. If no single point is predicted to belong to a particular face, then 32 points with coordinates (0,0,0) are used instead.

Because it is not effective to flatten feature vectors originating from 3D point clouds, and apply fully connected layers in order to predict classes or regress values from it, we process each subset individually by a PointNet module (Qi et al., 2017a). Here, four PointNet modules are used, so that each module can specialize on a specific face orientation. In order to enable the PointNet modules to also extract geometric features in addition to the ones resulting from the ConvPoint feature extraction component, the 3D point coordinates are concatenated with the respective feature vectors. Each PointNet module consists of a shared multilayer perceptron with 64, 128, and 256 output channels, an average pooling layer, and three fully connected layers that output 128, 64, and S+2 channels, where S is the number of slope classes. Besides the classification of the slope, the PointNet module regresses one slope correction value, and one binary classification score that determines the existence of the roof face by a probability value when the sigmoid (logistic) function is applied. The binary classification value acts like an objectness score that is commonly found in object detection.

3.3 Loss Function

The neural network produces four kinds of outputs: the point-wise roof face segmentation, four roof slope angle classes, four roof slope correction values, and four binary classifications that indicate the presence of the four roof faces, respectively. Thus, a multi-task loss function is used, which consists of four parts that are simply added up. For the roof face segmentation and slope classification, the mean cross entropy loss is used, for the roof slope correction values the mean absolute error (L1) loss, and for the binary roof presence classification the binary cross entropy loss. Since the two loss values for a roof face slope (class and correction) are only meaningful if the corresponding roof face does indeed exist, only the values of those faces are included in the loss function for which the binary values in the ground truth gives the value 1, thus indicating the presence of the roof faces.

4. 3D BUILDING MODEL CONSTRUCTION

The goal of this work is to construct geometric 3D building models from the neural network outputs, the given footprints, and under the previously postulated assumptions. For this purpose, half-space modelling is used, where a simple convex polyhedron is described as the Boolean intersection of multiple half-planes. See, e.g., (Kada et al., 2017) for a more thorough explanation of how half-space modelling can be applied for a geometric 3D modelling of buildings. At least five half-spaces are defined for the main body, one for the building base, and four or more for the building sides (the facades). The half-plane for the building base is directed straight downwards and positioned at the lowest point of the building. This lowest point is currently determined from the input 3D point cloud, which sometimes leads to unnaturally low buildings as the point cloud often does not contain further façade or ground points besides the roof points. However, this problem is easily solvable by providing a base elevation of the building or by taking further 3D points of the building's surrounding into consideration, and it does not change the principle applicability and correctness of the approach. The other four half-planes for the building sides are oriented perpendicular to the four line segments that form the rectangular 2D building footprint and the base half-plane, and are positioned accordingly. To give the 3D building model more details, half-planes from the convex hull of the real footprint can be used instead of the generalized rectangular footprint, which is shown in the results. Footprints with concave shapes are also possible, but then require another processing step as shown in (Kada at al., 2017).

For the building roof, up to four half-planes are generated from the neural network's outputs. For each possible roof face, a half-plane is only constructed if the neural network predicts its presence with a probability of at least 50%. The orientation of a half-plane for a roof face is determined by the rotation of the respective façade half-plane around the line of the line segment of the rectangular footprint by the predicted slope angle. And finally, the positions of these roof half-planes are calculated from the mean coordinates of all input points that are predicted by the roof face segmentation to belong to these faces. The intersection of the half-planes forms the geometric 3D building model in half-space representation, which can be converted into a boundary representation for visualization or further uses.

5. EXPERIMENTS AND RESULTS

5.1 Dataset

The proposed neural network is trained and evaluated with the RoofN3D dataset (Wichmann et al., 2019), which consists of around 118,000 simple buildings with a single saddleback, hip, or pyramid roof each. For each building, the dataset includes a variety of information, including a 3D point cloud for the entire object, a 2D footprint polygon, a roof type, a segmentation of points into planar roof faces, their plane equations, and much more. Out of all contained buildings, only those were used that contain 100 to 350 points, which resulted in 67,383 buildings. Although the neural network is designed to take any number of input points, the rationale for limiting the number of points to this range is on the one hand to avoid selecting duplicate points in the first sampling layer of the ConvPoint module, and on the other hand to make sure that sufficient points are sampled from the often small hip faces that might otherwise get missed if the ratio between sampled and input points is very low.

Before taken as input to the network, the 3D point cloud is brought into a local coordinate system by translating its center to the origin of this system. From the provided RoofN3D data, the class labels for roof face segmentation is derived according to the orientation of the roof faces in which they are located. Since we use random rotations around the upright axis for data augmentation, the class labels are adapted accordingly. No scaling is performed. From the half-plane equations of the roof faces provided by RoofN3D, the slope angle classes and their corrections are calculated for the existing faces, and the binary values for their presence taken. The possible range of 90° in the roof face slopes is divided into 18 classes of 5° angles, where the class always represents the angle in the middle of this range (e.g. 2.5° for the first angle class), which results in correction values in the interval [-2.5, 2.5). This seems a suitable compromise between the accuracy expressed by the resulting 18 angle classes per se, and the still quite small number of classes.

5.2 Network Training

The training of the neural network is performed with the Adam optimizer and a learning rate of 0.001. When using a random subset of 85% of the buildings for training, it takes around 40 epochs until no major improvements can be observed for the validation data. In the experiments, no signs of overfitting were observed. However, it cannot be completely ruled out that more epochs will not bring further improvements, since in particular the corrections of the roof slopes develop only very slowly.

In the experiments it was observed that using average pooling as the symmetric function in the PointNet modules gives more accurate results and leads to a more stable training process than max pooling. When using max pooling, large fluctuations in the evaluation metrics between epochs were observed and eventually led to a divergence of the training process.

5.3 Evaluation

The following quality metrics are reported for the evaluation of the four outputs of the proposed neural network: intersection over union (IoU) for roof face segmentation, overall accuracy (OA) for the roof slope classes, mean absolute error (MAE) for the roof slope corrections, and overall accuracy (OA) for the presence of roof faces. In addition, the resulting slope classes and corrections are transformed back to slope angles and the mean absolute error (MAE) is given compared to the original slope angles. Table 1 reports the results for two experiments: the first with a training data split of 85% for training, 10% for validation, and 5% for testing, and the second with a split of 95% for training and 5% for testing with the rationale to have a larger portion of the dataset for training. The data set was split in a stratified fashion, so that approximately the same proportion of all three roof types are represented in each split. The validation split was used to determine the number of epochs needed to train the neural network and to experiment on the neural networks hyperparameters. In both experiments, the neural network was trained for 40 epochs. The achieved scores can be considered quite high, especially considering that the ground truth data was generated automatically and does not necessarily reflect the real situation.

	85/10/5	95/0/5
Segmentation (IoU)	0.8027	0.8023
Slope classes (OA)	0.7653	0.7753
Slope correction (MAE)	1.1227	1.1061
Slope angle (MAE)	1.7429	1.8065
Face presence (OA)	0.9573	0.9575

Table 1. Scores for different training/validation/testing splits.

For visual quality inspection, a random subset of 625 buildings from the testing data split were reconstructed as 3D models. The resulting roof face segmentation and the 3D building models are depicted in Figure 2. Both types of outputs show convincing and visually accurate results for most of the objects. This is

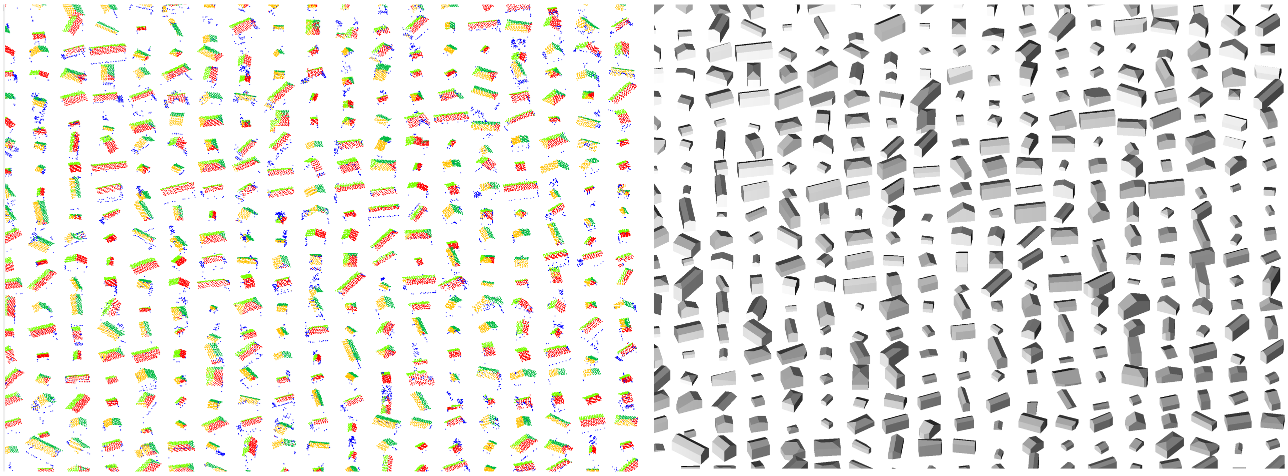


Figure 2. Results from roof face segmentation (left) and the reconstructed 3D building models (right).

especially true for building objects that can be clearly categorized into one of the three roof classes and where the existing faces are represented by seemingly sufficient points in the input 3D point cloud. Problematic are mostly buildings that do not that clearly belong to one of these roof shapes or where individual faces were not recognized. A further close-up view of the output is given in Figure 3.

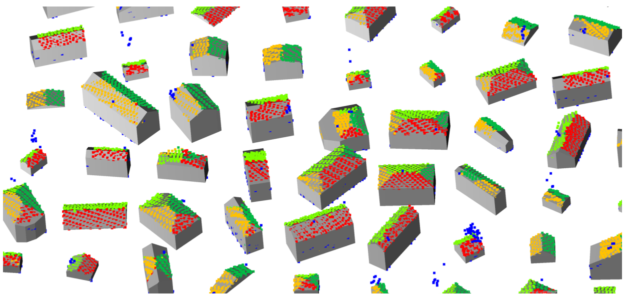


Figure 3. Examples showing results of roof face segmentation together with the resulting 3D building models.

The mean absolute error of the predicted roof face slope angles of 1.8° seems to be sufficiently accurate so that the geometric 3D building models fit quite well with the 3D point cloud. An example for a reconstructed building model with saddleback roof is depicted in Figure 4, which shows a very good fit of the two roof faces to the 3D input points.

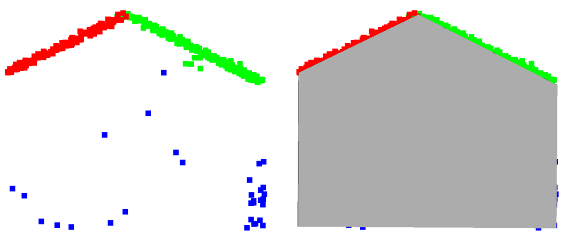
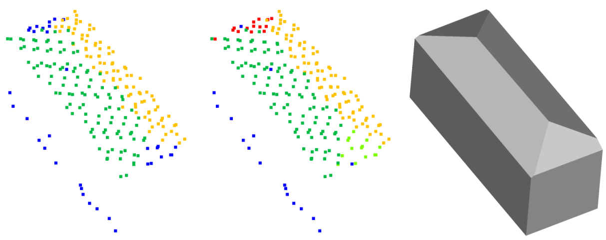


Figure 4. Side view of the roof face segmentation of a building with saddleback roof, and overlaid with the final 3D model.

It is interesting to note that the roof face segmentation seems to generalize rather well. Roof points that are incorrectly labeled in the training data, and roof faces that are therefore missing, are still well recognized by the neural network. In the example of Figure 5, points of the hip faces on both sides of the roof are marked in the training data as (blue) outliers, but the neural network correctly detects the two (red and light green) faces and can also calculate the slopes correctly, so that a matching 3D building model with a hip roof is reconstructed.



(a) (b) (c)

Figure 5. (a) Roof face segmentation of training data, (b) predicted segmentation, (c) reconstructed 3D building model.

6. CONCLUSION

In this paper, a neural network architecture is proposed capable of predicting geometric roof slope information and the presence of roof faces from 3D point clouds that allow the construction of 3D building models in half-space representation. Predicted half-planes seem to generally fit very accurately if a large enough number of points are found to belong to a roof face. It is shown that with such a reconstruction approach, primitives of simple buildings with saddleback, hip, and pyramid roofs can be predicted. Integrated into an object detection network that identifies and localizes building parts, the proposed module could be a stepping stone towards designing a Deep Learning based 3D building reconstruction architecture for objects with more complex shapes.

The general reconstruction approach and the proposed neural network architecture are not limited to the three roof types that were used in the experiments. It would be quite easy to include a flat roof point class to, e.g., also support flat and mansard roof types, and to train and predict other roof shapes having these five roof faces like pent, half-hip, broken-hip, etc. The only limitation is that the shape of the roof as well as the final building model must be convex. It is rather the lack of training and particularly testing data that limits further studies, since the RoofN3D dataset only contains buildings with these three roof types.

REFERENCES

Alidoost, F., Arefi, H., Tombari, F., 2019. 2D Image-to-3D Model: Knowledge-Based 3D Building Reconstruction (3DBR) Using Single Aerial Images and Convolutional Neural Networks (CNNs). *Remote Sensing*, 11(19), 2219.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A., 2015. Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information*, 4(4), 2842-2889.

Boulch, A., 2020. ConvPoint: Continuous Convolutions for Point Cloud Processing. *Computers & Graphics*, 88, 24-34.

Buyukdemircioglu, M., Kocaman, S., Kada, M., 2022. Deep Learning for 3D Building Reconstruction: A Review. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022, 359-366.*

Haala, N., Brenner, C., Statter, C., 1998. An Integrated System for Urban Model Generation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXII (2), 96-103.

Haala, N., Kada, M., 2010. An Update on Automatic 3D Building Reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65 (6), 570-580.

Henn, A., Gröger, G., Stroh, V., Plümer, L., 2013. Model Driven Reconstruction of Roofs from Sparse LiDAR Point Clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 76*, 17-29.

Kada, M., McKinley, L., 2009. 3D Building Reconstruction from LiDAR Based on a Cell Decomposition Approach. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-3/W4, 29-43.

Kada, M., Wichmann, A., Manzke, N., Filippovska, Y., 2017. A Data Model for the Interactive Construction and Correction of 3D Building Geometry Based on Planar Half-Spaces. *Abdul-Rahman, A. (eds.) Advances in 3D Geoinformation, Lecture Notes in Geoinformation and Cartography*. Springer, Cham., 461-476.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. PointCNN: Convolution On χ-Transformed Points. *Advances in Neural Information Processing Systems*, 31.

Maas, H.G., Vosselman, G., 1999. Two Algorithms for Extracting Building Models from Raw Laser Altimetry Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2-3), 153-163.

Mahmud, J., Price, T., Bapat, A., & Frahm, J. M., 2020. Boundary-aware 3D Building Reconstruction from a Single Overhead Image. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 441-451.

Nan, L., Wonka, P., 2017. PolyFit: Polygonal Surface Reconstruction from Point Clouds. *Proceedings of the IEEE International Conference on Computer Vision 2017*, 2353-2361.

Nguatem, W., Drauschke, M., Mayer, H., 2013. Roof Reconstruction from Point Clouds using Importance Sampling. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3/W3, 73-78.

Oude Elberink, S., Vosselman, G., 2009. Building Reconstruction by Target Based Graph Matching on Incomplete Laser Data: Analysis and Limitations. *Sensors*, 9(8), 6101-6118.

Qi, C.R., Su, H., Mo., K., Guibas, L.J., 2017a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 77-85.

Qi, C.R., Yi, L., Su., H., Guibas, L.J., 2017b. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Advances in Neural Information Processing Systems*, 30.

Qi, C.R., Litany, O., He, K., Guibas, L.J., 2019. Deep Hough Voting for 3D Object Detection in Point Clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9277-9286.

Qian, Y., Zhang, H., Furukawa, Y., 2021. Roof-GAN: Learning to Generate Roof Geometry and Relations for Residential Houses. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2796-2805.

Thomas, H., Qi, C.R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. KPConv: Flexible and Deformable Convolution for Point Clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6411-6420.

Verma, V., Kumar, R., Hsu, S., 2006. 3D Building Detection and Modeling from Aerial LiDAR Data. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2213-2220.

Wichmann, A., Agoub, A., Schmidt, V., Kada, M., 2019, RoofN3D: A Database for 3D Building Reconstruction with Deep Learning. *Photogrammetric Engineering & Remote Sensing* 85(6), 435-443.

Wu, W., Qi, Z., Fuxin, L., 2019. PointConv: Deep Convolutional Networks on 3D Point Clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9621-9630.