DEEP LEARNING FOR SEMANTIC 3D CITY MODEL EXTENSION: MODELING ROOF SUPERSTRUCTURES USING AERIAL IMAGES FOR SOLAR POTENTIAL ANALYSIS

S. Krapf[1,*], B. Willenborg[2], K. Knoll[1], M. Bruhse[2], T. H. Kolbe[2]

[1] Institute of Automotive Technology, Technical University of Munich (TUM), Munich, Germany
sebastian.krapf@tum.de
[2] Chair of Geoinformatics, Technical University of Munich (TUM), Munich, Germany
(b.willenborg, thomas.kolbe)@tum.de

Commission IV, WG IV/9

KEY WORDS: Semantic 3D City Models, CityGML, roof superstructure reconstruction, increasing level of detail, deep learning, solar potential

ABSTRACT:

On a global scale, semantic 3D city models with Level of Detail 2 become more and more available. Automated generation of higher Level of Detail models is an active field of research, but low coverage of dense LiDAR or photogrammetric point clouds is a barrier. Therefore, this paper presents a novel approach for enriching semantic 3D city models with roof superstructures extracted from aerial images using deep learning. The method maps and classifies superstructures in 2D and subsequently transforms them to 3D. Furthermore, we examine the benefit of the enriched model for solar potential analysis. The accuracy of solar potential analysis is improved by avoiding invalid simplifications of slope, shadow and panel placement. The enriched model reduces overestimation of accumulated solar potential by around 20 % compared to an analysis based on aerial images only. The novel method contributes to increasing the availability of Level of Detail 3 models for larger areas, while posing further research opportunities.

1. INTRODUCTION

Semantic 3D city models describe the geometric, visual, and thematic aspects of the most important entities of cities by decomposing and classifying them according to a semantic data model. CityGML is an open data model and encoding standard for the representation and exchange of semantic 3D city models (Kolbe, 2009; Open Geospatial Consortium, 2021). As of today, city models according to the CityGML standard are widely adopted for various use cases and application scenarios ranging from purely visualization tasks to complex analytic systems and simulations (Biljecki et al., 2015). 3D city model data has become available nationwide in Level of Detail 1 (LoD1) and Level of Detail 2 (LoD2) in big parts of Europe, as well as a number of countries in Asia and the Middle East, where mapping authorities create, update and maintain the models. Globally, there is also a growing number of CityGML models, of which more and more are offered as open data (Wysocki et al., 2022).

Currently, only LoD1 and LoD2 models can be generated with a high degree of automation from LiDAR or dense image matching point clouds and building footprints, e. g. from official cadastre (Kada and McKinley, 2009; Haala and Kada, 2010; Mc-Clune et al., 2016). Architectural details such as dormers, chimneys, or other roof superstructures (RSS) are usually not included in the models. This lack of detail is a limiting factor for many applications, as for instance, the estimation of solar energy potentials on building roofs. Even though the quality and availability of input data such as point clouds and aerial images is increasing and the 3D reconstruction of buildings is an active field of research, (see Section 2), to this date the problem of automatically deriving semantically classified RSS for semantic 3D city models remains unsolved. In particular, the

combination of geometry and semantics is rarely seen, which is a key feature making semantic 3D city models an attractive data integration platform for a wide range of simulations and urban analytic applications, compared to other 3D building representations such as 3D mesh models.

In recent years, researchers successfully applied deep learning-based methods for automatic extraction of semantic, georeferenced information from aerial images. Building on these advances, this article explores a novel approach of enriching semantic 3D city models with semantically classified RSS, making use of the high degree of detail in high resolution aerial images as well as deep learning methods for automatized mapping. To investigate the effect of increased information in semantic 3D city models on applications, the results are evaluated and discussed on the example of solar potential analysis.

2. RELATED WORK

2.1 Building and superstructure reconstruction

Building reconstruction is an active field of research in GIS, computer vision, and photogrammetry. Conventional building reconstruction methods can be classified into data-driven and model-driven approaches. Data driven methods rely on the fitting of individuals planes to reconstruct buildings, while model-driven approaches operate by estimating the parameters of pre-defined primitives Zhang et al. (2021). Due to rapid developments in airborne laser scanning and stereo photogrammetry, a variety of algorithms have been proposed to generate 3D building models up to LoD2 from point clouds. However, according to Wichmann and Kada (2016), few research has dealt with the automated reconstruction of roof superstructures in semantic 3D city models. Stilla and Jurkiewicz (1999) present an early

---

* Corresponding author

reconstruction approach based on histograms from Digital Elevation Model (DEM) to detect flat superstructures on flat buildings. Bredif et al. (2007) also use DEM and apply a model-based approach to detect volumetric superstructures such as dormers and chimneys. Wichmann and Kada (2016) propose a method which uses superstructure symmetry to improve reconstruction results for low-density LiDAR data. Meixner et al. (2011) combine a Digital Surface Model (DSM) and orthophotos to map dormers, chimneys and other volumetric superstructures. The authors are able to detect flat superstructures (skylights) using a template matching approach. Zhang et al. (2021) introduced a reconstruction procedure to derive a CityGML LoD2 model with superstructures modeled as building installation objects from point clouds. For the reconstruction of complex roofs a hierarchical procedure is presented to reconstruct the major roof model and its superstructures sequentially based on primitive parameterization and recognition. However, the number of supported superstructure primitives is limited and flat superstructures are not dealt with.

In recent years, deep learning methods have experienced increasing research interest, due to their large potential of automated extraction of information from remote sensing data. Point cloud data does not represent small or flat objects, therefore, a number of researchers have used aerial images as an alternative source of input. For example, Yu et al. (2018) propose the DeepSolar framework to map solar panels in the USA. Mayer et al. (2020) apply the network to the German state North-Rhein-Westphalia. In a subsequent study, Mayer et al. (2022) combine the image recognition approach with semantic 3D city models to improve the estimation of existing solar capacity, indicating the potential synergies between the two data sources. Mainzer et al. (2017) and Nelson and Grubesic (2020) apply classic computer vision techniques to detect superstructures in aerial images Respectively, a study from Krapf et al. (2021) shows that deep learning can be used for roof superstructure detection. A major barrier for deep learning approaches is the requirement of labeled training data. Therefore, Krapf et al. (2022) published their data set for semantic segmentation for roof segments and roof superstructures.

2.2 Models for simulation of solar potential

To support policy makers and to inform the public, researchers have conducted rooftop solar potential analyses ranging from city-scale to country-scale. Highly accurate solar analysis requires a range of building and roof information such as geographic location, roof geometry, slope angle, azimuth angle and shadowing effects of surrounding topography, and is therefore an effective use case for evaluating the benefit of increasing model Level of Detail (LoD). The publications of Freitas et al. (2015); Gassar and Cha (2021) give a good overview on the existing approaches for modeling solar potential in an urban environment. Solar potential analysis can be structured according to its input data based on statistical data, geo-spatial data, aerial images or 3D data. 3D approaches use existing semantic 3D city models (Willenborg et al., 2018; Romero Rodríguez et al., 2017) or LiDAR data and are typically applied on city-scale, for example Lisbon (Brito et al., 2017). Although 3D data is becoming increasingly available, there is still no extensive coverage on global scale, especially in more rural areas. Therefore, researchers extrapolate the detailed analysis of regions with 3D data to larger scales. Gagnon et al. (2018) statistically extended an analysis of 128 US cities from Margolis et al. (2017) to the whole USA. Walch et al. (2020) apply machine

learning methods to expand the results from Geneva to Switzerland. Other publications combine LiDAR data and aerial images for automated extraction of roof segments to counteract low LiDAR point resolution (Martín-Jiménez et al., 2020). Due to higher spatial coverage of aerial images some publications propose solar potential analysis based on aerial images only, and capitalize on the emerging use of deep learning in the remote sensing field as described in Section 2.1. Lee et al. (2019) introduce a deep learning framework for extracting roof segments from images and assigning an azimuth class. Krapf et al. (2021) extend this idea by proposing a Convolutional Neural Network (CNN) for the task of roof superstructure segmentation, in addition to the CNN for roof segments.

2.3 Summary, Research Gap and Contributions

Reconstruction of roof superstructures remains an ongoing field of research. While LiDAR data is used to create the building information of semantic 3D city models, it is not dense enough to include small objects like chimneys and does not contain flat objects like windows or existing solar modules. Therefore, this paper explores a novel approach of extracting RSS from aerial images with deep learning and enriching 3D city models with this information. The proposed method uses the semantic information and structures of an existing LoD2 semantic 3D city models and extends the available information by integrating geometry and semantics of RSS. Solar potential analyses requires a high LoD for high accuracy. Therefore, we conduct a solar case study to evaluate the benefits and challenges of our method and enriched semantic 3D city models. The contribution of this paper are summarized as:

- A novel method for increasing semantic 3D city models LOD with roof superstructures extracted from aerial images using deep learning

- An evaluation of the viability of our approach based on solar potential analysis

- Accuracy improvements of solar potential analysis

Section 2 briefly introduced related work, derived the research gap and presented our contributions. The rest of the paper is structured as follows. Section 3 outlines the process of enriching semantic 3D city models with RSS in Section 3.4 and describes the steps of the solar potential analysis in Section 3.5. The case study and data sets are presented in Section 4. Then, Section 5 presents the results of the solar potential analysis based on aerial images only, semantic 3D city models only and the enriched semantic 3D city models. Benefits and limitations of the method are discussed in Section 6.

3. METHOD

The core idea of this work is to enrich 3D city models with geometries and semantic information gained from aerial images using deep learning. It is important to note that the method presented in the following section is being explored at the moment. The approach is currently only capable of reconstructing simplified RSS. However, the output can already be used to increase the LoD of semantic 3D city models for subsequent simulations and analysis applications, that do not necessarily require an exact representation.
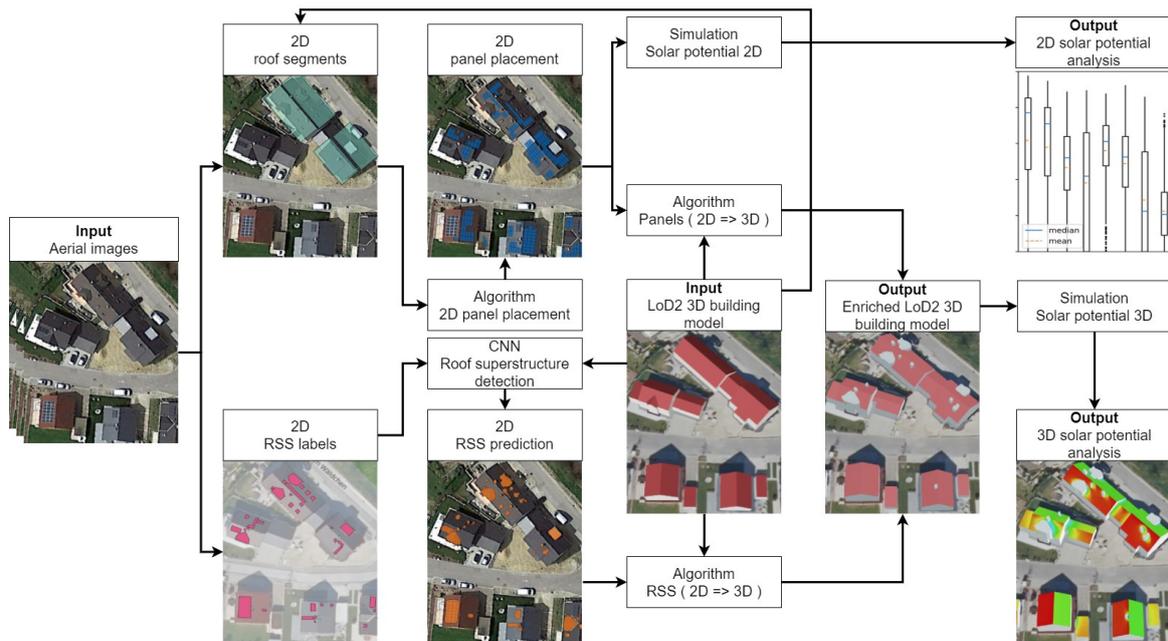
Figure 1. Overview on the input data, model components of bringing information from 2D to 3D. CNN = Convolutional neural network, RSS = Roof superstructures.

3.1 Overview on the proposed method

The core components of the proposed method and how they are connected to each other, are illustrated in Figure 1. The first part of the approach is based on 2D aerial images as input. Using two different deep learning networks, roof segments and RSS are detected and classified. This information is used to apply an algorithm for placing solar panels in 2D on the building roofs. In this paper, 2D solar potential analysis is performed as a comparison to the enriched semantic 3D city models (see Section 5.2). After panel placement, the 2D geometries are elevated to the third dimension and integrated into a semantic 3D city models of the same area, which is the second input of the method. Finally, a 3D solar potential analysis is performed with the enriched semantic 3D city models.

3.2 Deep learning for 2D roof superstructures

Krapf et al. (2021) describe the deep learning approach of extracting roof segments and RSS from aerial images. In contrast, this paper uses roof segment data from semantic 3D city models instead of predictions by a CNN. For the task of semantic segmentation of RSS from aerial images, we use a Feature Pyramid Network (FPN) as implemented by Yakubovskiy (2019) based on the design by Kirillov et al. (2019). The publication by Krapf et al. (2022) provides detailed information regarding the training and evaluation of the CNNs for superstructure detection as well as the respective training data set. While for this paper, a larger ResNet-152 backbone was trained, the prediction accuracy did not change significantly. The network was initialized with weights pretrained on the 2012 ILSVRC ImageNet dataset. The model and training parameters are listed in Table 1. The ground-truth dataset is described in Section 4.

The true positive predictions in the confusion matrix (Figure 2) show that the model performed well when detecting the classes relevant in the context of 3D RSS modeling: Dormers have a true positive rate of 0.59 and chimneys 0.60. Only solar modules were predicted more accurately (0.75), as shown in Figure 2.

Parameter	RSS CNN
Backbone	ResNet-152
Training Epochs	40
Learning Rate	0.000 1
Optimizer	Adam
Loss Function	Focal + weighted Jaccard
Activation	Sigmoid

Table 1. Model and training parameters of the used CNN.

3.3 Enriching 3D city models with roof superstructures

The predictions of 2D RSS (see Section 3.2) are used to create 3D RSS in a CityGML model. The images used are roof centered. Because of this, buildings that are close to one another will appear in multiple images. Consequently, the RSS on those roofs will be predicted multiple times by the CNN. To avoid generating duplicate RSS in the CityGML model, overlapping RSS of the same type are unified into one geometry. If RSS of differing types overlap, classes with more information value for the 3D model are preferred in the following order:
$$unknown < ladder < window < pvmodule < chimney < dormer$$
The predicted masks (raster data) are translated into vector data by creating polygons from connected regions of the same class using the GDAL library. The edges of these polygons follow the pixel boundaries exactly, which results in very dense geometries with many points. Because of this, the polygons are simplified using the Douglas-Peucker algorithm with a 0.1 m tolerance. The extracted polygons lack a z-coordinate. To generate a correct representation in the 3D model, their height values must be calculated. For this, the polygons' points are paired to a roof of the LoD2 model using intersection. Then, roof slope and orientation are derived from the roof's normal vector. With roof slope, orientation and one arbitrary point on the roof known, the z-values of the intersecting points can be calculate. In the case of flat RSS like solar modules, this z-value is then increased by 0.1 m to avoid clipping issues when viewing the 3D model. Volumetric RSS (chimney and dormer) are constructed by extruding the previously generated flat polygons along
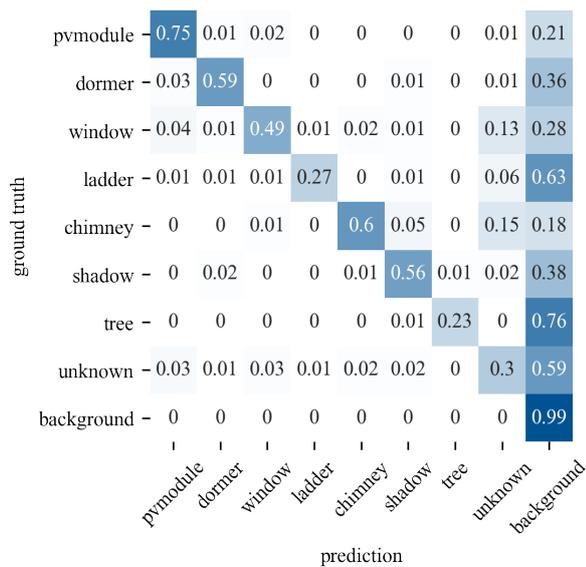
Figure 2. Confusion matrix for the RSS detection CNN.

the z-axis. For chimneys, this extrusion lasts until all points of the chimney's top polygon are at least $1\,\mathrm{m}$ above the roof surface. Dormers are extruded until all points of the top polygon are level with the highest point of the base polygon.

3.4 Automatic solar panel placement in 2D and 3D

To determine the solar potential of newly installed solar systems on roof area not obstructed by existing superstructures, a module placement algorithm is required. First the algorithm places modules on a 2D roof segment, then, the panels are transformed and integrated into the semantic 3D city model. Based on the roof segment polygon, modules are placed within the polygon in parallel to the roof segment's gutter starting at the polygon's origin (bottom-left corner). Module dimensions are transformed to represent the projection of the module on the plane. Panels that lie outside the roof segment or intersect with superstructures are deleted. The algorithm computes horizontal and vertical module orientation and chooses the configuration with a greater number of modules. On flat roofs, modules can be oriented south, east-west, or in alignment with the building. Spacing between rows is added and the modules are pitched at $30°$. Finally, the modules are transformed to 3D with the same method as described for the RSS in Section 3.3.

3.5 Solar potential estimation

This section describes the solar potential analysis based on aerial images and deep learning and the 3D-data based analysis. Both approaches function as reference cases to compare the benefits of combining semantic 3D city models with superstructures extracted from aerial images. Solar potential can be assessed in four different forms, as physical, geographic, technical or economic potential. In this paper, we compute geographic potential and define it as the sum of radiation on tilted plane at a specific location in one year, in $\frac{\mathrm{kW\,h}}{\mathrm{a}}$.

3.5.1 Calculation of solar potential in 2D based on aerial images and AI The solar potential analysis based on aerial images and deep learning is described in detail by Krapf et al. (2021). In this section, we summarize the central steps of the framework. First the radiation on horizontal plane is determined for one point in the entire study area using the Copernicus

Atmosphere Monitoring Service (ECMWF, 2021). Due to the small size of the study area, the variance of the radiation is negligible. Next, RSS are extracted from aerial images with a deep learning computer vision approach as described in Section 3.2. The framework identifies all buildings in the study area based on map data, e.g. OSM (OpenStreetMap contributors, 2017), and obtains an aerial image for each building with the roof at the center. The CNN outputs the RSS information as predictions masks which are subsequently transformed from raster to vector format. In the 2D case, the azimuth angle of the roof segments is classified into 16 azimuth classes (e.g. South, South-South-East, South-East, etc.) and one flat roof class. Aerial images do not contain height information, so the roof slope is approximated by statistical data with a mean of $37°$ and standard deviation of $15°$ as proposed by Mainzer et al. (2017). The shadowing effect is included as constant factor of $15\,\%$. With this information we calculate the geographic potential as radiation on tilted plane. It consists of direct, diffuse and ground-reflected radiation. We apply the model by Perez et al. (1990) to calculate the diffuse radiation component using pvlib (Holmgren et al., 2018). Section 3.4 describes the solar module placement.

3.5.2 Solar potential analysis with 3D city models The solar irradiation analysis tool presented by Willenborg et al. (2018) used in this work, computes the direct and diffuse solar irradiation and the SkyView-Factor (SVF) for roofs and facades considering shadowing effects of buildings, a DSM, and other features represented as CityGML objects, e.g. Vegetation, or Bridge objects. The influence of reflected radiation is neglected. For this work, the tool was extended to use the solar panels as calculation basis to support the enriched 3D city model derived with the method described in Section 3. The radiation on the tilted plane is computed for each solar module with a sampling point grid resolution of $0.15\,\mathrm{m}$. In addition to shadowing effects of surrounding buildings, the enriched semantic 3D city models integrates shading of RSS. To evaluate the benefits of the enhanced model for this use case, we conduct a case study in several different configurations, which is described in Section 4.

4. CASE STUDY

To our knowledge, the Roof Information Dataset (RID) by Krapf et al. (2022) is the only publicly available data set for semantic segmentation of RSS. In this paper, we use this RSS data and therefore, our case study area is limited to the data of RID. Krapf et al. (2022) selected the rural German town Wartenberg as area of interest because of its size of around $2\,000$ buildings and the available high resolution areal images $10\,\frac{\mathrm{cm}}{\mathrm{px}}$ with simultaneously challenging image quality with respect to contrast, image distortion and light conditions. Krapf et al. (2022) describe the labeled data set in detail. The aerial images are from the year 2018 and were retrieved from the Google Maps Static API. Based on the images, the annotators manually labeled $12\,359$ superstructures on $1\,880$ buildings. For each annotated building one image is downloaded with the building outline's centroid at the center of the image. Hence, images overlap and cover a total area of $4.9\,\mathrm{km}^2$ with a uniquely covered area of $1.5\,\mathrm{km}^2$. Superstructures are classified into eight classes, *pvmodule* for existing solar panels, *dormer*, *window*, *ladder*, *chimney*, *shadow*, *tree*, and *unknown* to account for other unclassified superstructures. The superstructure classes *shadow* and *tree* do not prevent the placement of solar panels and are therefore excluded from potential assessment. Roof architectures and RSS are subject to great variance resulting in high class imbalance for both,
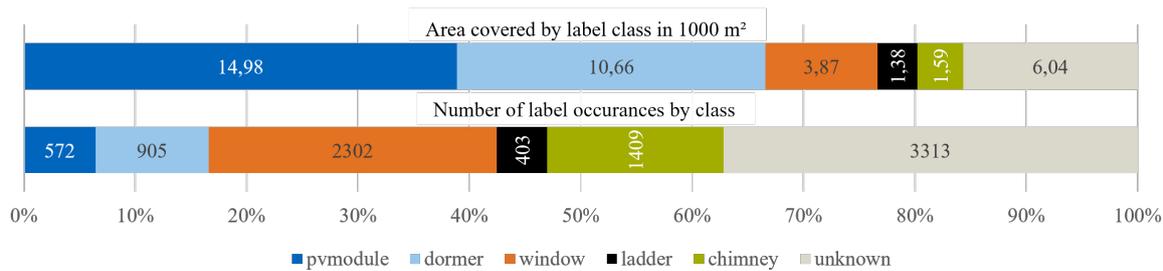
Figure 3. Class balance of the superstructure dataset.

number of labels per class and area covered per label class, as illustrated by Figure 3. While there are only 572 labeled polygons of class *pvmodule*, they cover approximately $15\,000\,\text{m}^2$. In contrast, the largest number of annotations is from *unknown*, covering only an area of around $6\,000\,\text{m}^2$. In total, the projected area of superstructures is $38\,500\,\text{m}^2$ making up $12.1\,\%$ of the roof area projection. To evaluate the presented method for enriching 3D city models with RSS from aerial images, solar potential analysis serves as an exemplary application. To this end, we calculate the solar potential in multiple configurations, explained in more detail in Section 5. Solar potential analysis based on aerial images, only, and based on 3D city model without RSS represent two baseline cases for comparison with the enriched model.

5. RESULTS

This section presents the results of modeling RSS predicted from aerial images in semantic 3D city models for the study area. Predicted RSS and solar panels are incorporated for the entire study area, but we report the solar potential results only for the buildings in the test set, to avoid inflated prediction results from the training and validation set. First, we discuss the results qualitatively in Section 5.1. Then, we quantify the overall solar potential in Section 5.2 to investigate the larger scale. Figure 4 shows the enriched semantic 3D city models for the study area and displays the town center. This visualization shows the ground truth labels of RSS instead of the predictions. On the left of the figure, solar modules are automatically placed on available roof areas while the right part illustrates the roof segments without superstructures only.

5.1 Qualitative results

The following qualitative analysis is conducted using an excerpt of five buildings within the test set as example (Figure 5). Figure 5a shows the Google Maps aerial image. The selected five buildings have roofs which are partly occupied by dormers, chimneys, ladders windows and existing solar modules. The next Figure 5b contains the 2D results, i.e. CNN prediction and module placement, which are used to calculate the baseline 2D solar potential. Figure 5c visualizes roof segments and their solar potential solely based on semantic 3D city models and the solar analysis described in Section 3.5.2. Each building has one segment with high solar potential (orange or red) due to the south-west or south-east orientation and a corresponding north-facing opponent with lower solar potential. The semantic 3D city models base model contains already one dormer on the left-most building. However, other dormers are not yet included and are only visible in Figure 5d which appends ground truth labels of the volumetric superstructures chimney and dormer. The

shadowing effect of these additional structures becomes apparent as the solar potential decreases in close proximity. Next, on the top right, Figure 5e portrays predictions as comparison to ground truth labels. The network maps all volumetric superstructures contained in the ground truth, but misclassifies two windows and a ladder as chimneys leading to additional volumetric superstructures on the three buildings in the upper half. Figure 5f shows an ortho-photo. While the Google image for this extract is almost orthogonal as well, the image distortion is high in some sections in the study area. Next, the panel placement is modeled in 3D (Figure 5g). This scenario uses ground truth data (segments and RSS) based on the ortho-photo and the placement algorithm was allowed to plan modules also on the area of existing panels. On the left segment of the bottom left building, the algorithm placed 20 modules, whereas in reality, 21 modules have been placed, but the modules appear to be slightly smaller in the aerial image. The effect of missing roof overhang modeling in semantic 3D city models can be observed by comparing Figure 5g and Figure 5h. The latter image's panels are placed using manually labeled roof segments which are based on Google Maps images. Therefore, the roof overhang is considered in contrast to all other 3D scenarios, where panels outside the semantic 3D city models's roof segments are discarded. The additional roof space leads to an increased number of modules and a higher solar potential. For example, 23 as opposed to 16 panels are assigned to the right segment of the bottom-left building. However, due to the naive placement and imprecise transformation between 2D and 3D, some buildings display a higher number of modules in the scenario without roof overhang, as can be seen on the south-facing segment of the top-most building. Figure 5i examines the effect of shade on solar panels. Especially on the top-most building, some of the south-facing panels are subject to shadowing and consequently, lower energy generation. In addition, the drawback of the simple modeling approach can be observed with respect to the dormer on the left-most building, which casts an intense shadow, even though the two small tilted planes actually block the radiation marginally (orange color in Figure 5h). Finally, Figure 5j displays the resulting enriched semantic 3D city models including RSS predicted by the CNN and solar panels. This scenario represents a more realistic model for solar potential assessment in comparison to the 2D and semantic 3D city models only approaches due to four aspects. Firstly, in comparison to using the whole roof segment area for solar potential calculation (Figure 5c), the radiation on solar panels leads to a decrease of the utilized roof area due to placement constraints. In addition, RSS induce shading losses, but more importantly, reduce the available roof area significantly (Figure 5i). Solar panels and RSS represent information which can be extracted from aerial images using deep learning to improve the analysis based on semantic 3D city models. Inversely, semantic 3D city models provide information for the 2D solar potential analysis

Figure 4. Study area in Wartenberg: On the left panels with RSS and on the right solar potential on complete roof segments are shown.

which lacks slope information and shading effects, also from the surrounding terrain. Roof slope is a critical value, because it affects the solar radiation on the solar panel. Moreover, the effect of slope on solar potential through the panel placement can be even more severe, especially for large tilt angles, because panels are planned using the projected area on horizontal plane. For example, considering two roof segments with the same horizontal projection, the roof segment tilted by $70°$ is more than twice as large as the segment with $37°$ slope. Finally, surrounding buildings or RSS can block solar radiation and reduce energy generation. Figure 5d underlines, that the shading effect can vary strongly between roof segments. In particular, smaller adjoining auxiliary buildings such as garages are attractive roof segments for residential solar systems, but are often subject to higher degree of shading.

5.2 Total potential and segment specific potential

To investigate the difference between 2D, aerial image-based, semantic 3D city models-based and the combined, enriched semantic 3D city models, we calculate the geographical potential for the study area. We focus on the test set due to over-fitting of the CNN on training and validation set. We determine the geographical potential for each roof segment and then compute the sum to derive the total potential. Figure 6 visualizes the results as violin plots of specific geographical potential in $\frac{\text{kW h}}{\text{m}^2 \text{a}}$, i.e. the potential per segment normalized by the segments area. Results are presented for five configurations, referred to as C1 to C5 from left to right. They can be found as close-up view in Figures 5b, 5c, 5e, 5h and 5i, respectively. Segments without panels are excluded from the violin plots C1, C4 and C5 reducing the number of segments by 542 (41%). C1 constitutes the 2D baseline using slope values from the LoD2 model and leads to a total geographic potential of $18.32\,\frac{\text{GW h}}{\text{a}}$. The maximum specific geographic potential is about $1\,200\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$, while the study area's theoretical maximum is about $1\,400\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$ for a south-facing roof segment (azimuth $4°$) and a slope of $37°$. Most segments have specific potential of less than $800\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$. In contrast, configurations C2 and C3 display a median specific potential of more than $800\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$ and total potential of $45.30\,\frac{\text{GW h}}{\text{a}}$ and $42.62\,\frac{\text{GW h}}{\text{a}}$. The difference between C2 and C3 stems from shading effects on roof segments and amounts to $2.68\,\frac{\text{GW h}}{\text{a}}$ or 5.91% of C2. Furthermore, the violin plots of C2 and C3 show small peaks around $1\,100\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$ and $700\,\frac{\text{kW h}}{\text{m}^2 \text{a}}$ indicating the typical south-north roof orientation. The effect of panel placement

can be observed by comparing C2 and C4. The geographical potential's decrease is $29.72\,\frac{\text{GW h}}{\text{a}}$ or 65.61% from C4 to C2, highlighting the major difference between roof segment area and total solar module area. The influence of RSS shading on solar modules is lower than on the entire segment, as the potential is only reduced by $0.3\,\frac{\text{GW h}}{\text{a}}$ or 1.93% of C4 when considering RSS in configuration C5. However, this resembles the effect on geographical potential where modules are examined individually. In reality, shading on one module can reduce the energy generation of the entire PV system due to series connection of modules. The impact of shading due to surrounding buildings becomes visible through comparison of C1 and C4, i.e. the 2D case and the 3D case without RSS. The shadow of buildings causes potential reduction of around $2.76\,\frac{\text{GW h}}{\text{a}}$. C1 and C4 assume the same number of modules and slope value for the segments, excluding the influence of different number of modules from the analysis. We also calculated the geographical potential of the 2D case with random slope values. However, we found only a small difference of $0.02\,\frac{\text{GW h}}{\text{a}}$ and slightly more evenly distributed specific potential due to the random slope values creation using a standard variance.

6. DISCUSSION

This paper explores enriching semantic 3D city models with RSS information extracted from aerial images using deep learning. The enrichment of existing LoD2 models is preferable to a complete recreation, because the existing models contain manifold semantic information, which cannot be transferred easily. While first results indicate great potential, the presented approach reveals several limitations. First of all, the study area is a rural German town and scalability requires annotated RSS from larger, more diverse regions. Furthermore, 3D objects are modeled with great simplicity. A model-driven approach could apply primitives according to RSS class type, but class types are not detailed enough to model different types of dormers (e.g. hip vs. shed) or chimneys (e.g. squared vs. round). Another challenge is the geographical and temporal mismatch of aerial image data and semantic 3D city models as well as image distortion which can lead to erroneous locations of transformed objects or missing newly constructed buildings. Further types of artifacts can arise from the CNN's limited prediction quality. Additionally, a downside of the presented results is the infeasible validation due to lack of 3D ground truth data of
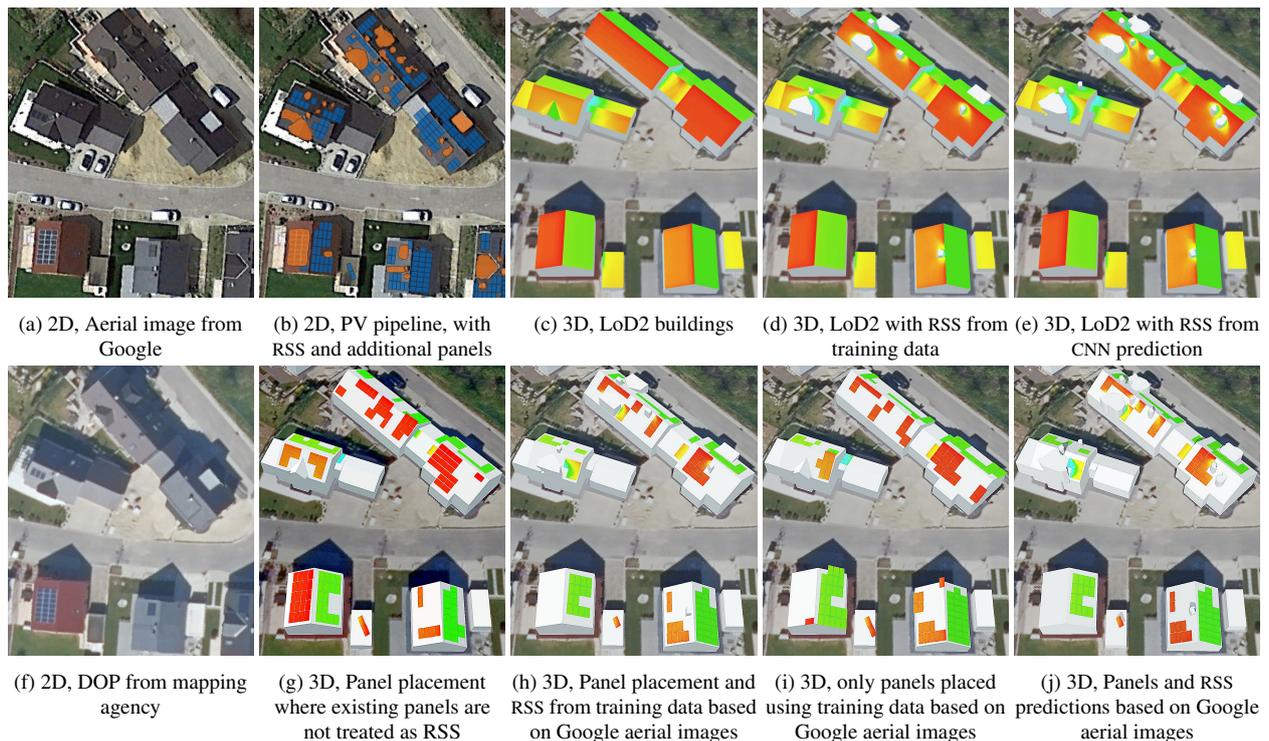
(a) 2D, Aerial image from Google

(b) 2D, PV pipeline, with RSS and additional panels

(c) 3D, LoD2 buildings

(d) 3D, LoD2 with RSS from training data

(e) 3D, LoD2 with RSS from CNN prediction

(f) 2D, DOP from mapping agency

(g) 3D, Panel placement where existing panels are not treated as RSS

(h) 3D, Panel placement and RSS from training data based on Google aerial images

(i) 3D, only panels placed using training data based on Google aerial images

(j) 3D, Panels and RSS predictions based on Google aerial images

Figure 5. Overview in the input datasets and experiments for the case study.
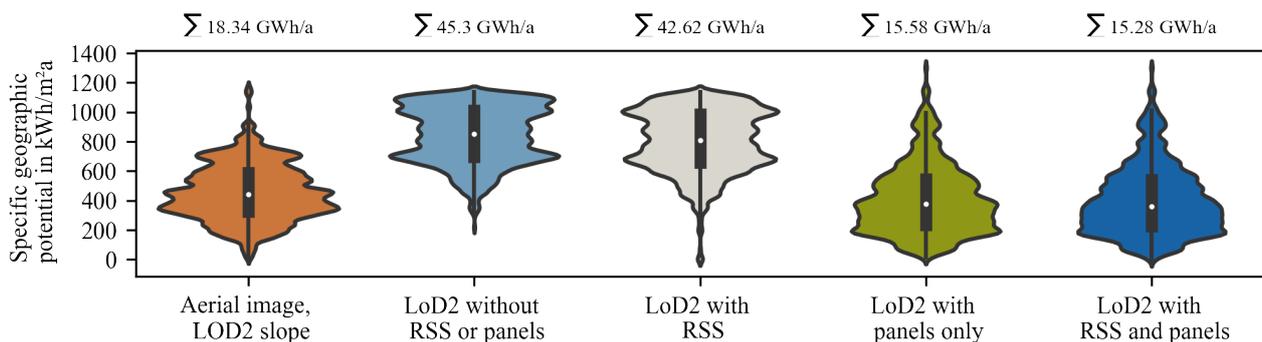
Figure 6. Specific geographical potential of roof segments in the study area.

RSS. Hence, results were only verified by visual inspection and quantitative validation remains an open point for future work. In our solar potential analysis, no solar panels were placed on more than 40 % of roof segments, which can partly be traced back to the different data sources. In Figure 5 we illustrated the option of keeping or dismissing roof overhangs, but either option poses unresolved challenges. Finally, with regard to solar potential analysis, the effect of shaded modules is simplified. In reality, a shaded module impacts other modules connected in series, which needs to be considered to quantify the real impact of RSS's shadow on solar potential.

7. OUTLOOK

Automated generation of Level of Detail 3 (LoD3) models is an active field of research, but low coverage of dense LiDAR or photogrammetric point clouds is a barrier. Therefore, this paper presented a novel approach for enriching semantic 3D city models with RSS extracted from aerial images with deep learning. We described the process of mapping and classifying

RSS in 2D and subsequent transformation to 3D. The benefit of increased information was examined based on solar potential analysis. Results showed the viability of the approach and revealed challenges and research opportunities. Furthermore, the enriched model improves solar potential analysis by avoiding invalid simplifications of slope, shadow and panel placement. A comparative analysis in a small study area exhibited that solar potential is overestimated by around 20 % compared to the 2D baseline case. Hence, the presented method contributes to increase the availability of LoD3 models on larger scale, while posing further research opportunities. Future work should tackle three major aspects: First, the scalability of RSS with deep learning needs to be expanded. Larger data sets including urban areas and different geographic regions as well as additional RSS classes are required. Second, modeling of RSS in semantic 3D city models should be improved. For example, a model-driven approach can be investigated based on existing RSS semantics. Lastly, the effect of LoD3 on solar potential should be examined in more detail, potentially requiring more advanced modeling of solar systems under the influence

of shadow and temperature.

REFERENCES

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.

Bredif, M., Boldo, D., Pierrot-Deseilligny, M., Maitre, H., 2007. 3d building reconstruction with parametric roof superstructures. *2007 IEEE International Conference on Image Processing*, IEEE, II – 537–II – 540.

Brito, M. C., Freitas, S., Guimarães, S., Catita, C., Redweik, P., 2017. The importance of facades for the solar PV potential of a Mediterranean city using LiDAR data. *Renewable Energy*, 111, 85–94.

ECMWF, 2021. Copernicus Atmosphere Monitoring Service. https://atmosphere.copernicus.eu/.

Freitas, S., Catita, C., Redweik, P., Brito, M. C., 2015. Modelling solar potential in the urban environment: State-of-the-art review. *Renewable and Sustainable Energy Reviews*, 41, 915–931.

Gagnon, P., Margolis, R., Melius, J., Phillips, C., Elmore, R., 2018. Estimating rooftop solar technical potential across the US using a combination of GIS-based methods, lidar data, and statistical modeling. *Environmental Research Letters*, 13(2), 024027.

Gassar, A. A. A., Cha, S. H., 2021. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. *Applied Energy*, 291, 116817.

Haala, N., Kada, M., 2010. An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 570-580. ISPRS Centenary Celebration Issue.

Holmgren, W. F., Hansen, C. W., Mikofski, M. A., 2018. pvlib python: a python package for modeling solar energy systems. *Journal of Open Source Software*, 3(29), 884.

Kada, M., McKinley, L., 2009. 3D Building Reconstruction from Lidar Based on a Cell Decomposition Approach. U. Stilla, F. Rottensteiner, N. Paparoditis (eds), *CMRT09*, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Citeseer, 47–52.

Kirillov, A., Girshick, R., He, K., Dollár, P., 2019. Panoptic feature pyramid networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6399–6408.

Kolbe, T. H., 2009. Representing and Exchanging 3D City Models with CityGML. J. Lee, S. Zlatanova (eds), *3D Geo-Information Sciences*, Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, 15–31.

Krapf, S., Bogenrieder, L., Netzler, F., Balke, G., Lienkamp, M., 2022. RID – Roof Information Dataset for Computer Vision-Based Photovoltaic Potential Assessment. *Under final review*.

Krapf, S., Kemmerzell, N., Khawaja Haseeb Uddin, S., Hack Vázquez, M., Netzler, F., Lienkamp, M., 2021. Towards Scalable Economic Photovoltaic Potential Analysis Using Aerial Images and Deep Learning. *Energies*, 14(13), 3800.

Lee, S., Iyengar, S., Feng, M., Shenoy, P., Maji, S., 2019. Deeproof: A data-driven approach for solar potential estimation using rooftop imagery. A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, G. Karypis (eds), *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, 2105–2113.

Mainzer, K., Killinger, S., McKenna, R., Fichtner, W., 2017. Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques. *Solar Energy*, 155, 561–573.

Margolis, R., Gagnon, P., Melius, J., Phillips, C., Elmore, R., 2017. Using GIS-based methods and lidar data to estimate rooftop solar technical potential in US cities. *Environmental Research Letters*, 12(7), 074013.

Martín-Jiménez, J., Del Pozo, S., Sánchez-Aparicio, M., Lagüela, S., 2020. Multi-scale roof characterization from LiDAR data and aerial orthoimagery: Automatic computation of building photovoltaic capacity. *Automation in Construction*, 109(12), 102965.

Mayer, K., Rausch, B., Arlt, M.-L., Gust, G., Wang, Z., Neumann, D., Rajagopal, R., 2022. 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Applied Energy*, 310(12), 118469.

Mayer, K., Wang, Z., Arlt, M.-L., Neumann, D., Rajagopal, R., 2020. Deepsolar for germany: A deep learning framework for pv system mapping from aerial imagery. *SEST' 20*, IEEE, Piscataway, NJ.

McClune, A. P., Mills, J. P., Miller, P. E., Holland, D. A., 2016. Automatic 3D Building Reconstruction from a Dense Image Matching Dataset. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3, 641–648.

Meixner, P., Leberl, F., Bredif, M., 2011. 3d roof details by 3d aerial vision. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, IEEE, 212–218.

Nelson, J. R., Grubesic, T. H., 2020. The use of LiDAR versus unmanned aerial systems (UAS) to assess rooftop solar energy potential. *Sustainable Cities and Society*, 61(1), 102353.

Open Geospatial Consortium, 2021. OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard v3.0.

OpenStreetMap contributors, 2017. Openstreetmap.

Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*, 44(5), 271–289.

Romero Rodríguez, L., Duminil, E., Sánchez Ramos, J., Eicker, U., 2017. Assessment of the photovoltaic potential at urban level based on 3D city models: A case study and new methodological approach. *Solar Energy*, 146, 264–275.

Stilla, U., Jurkiewicz, K., 1999. Automatic reconstruction of roofs from maps and elevation data.

Walch, A., Castello, R., Mohajeri, N., Scartezzini, J.-L., 2020. Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. *Applied Energy*, 262, 114404.

Wichmann, A., Kada, M., 2016. Joint Simultaneous Reconstruction of Regularized Building Superstructures from Low-Density LiDAR data using ICP. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, III-3, 371–378.

Willenborg, B., Sindram, M., Kolbe, T. H., 2018. Applications of 3D City Models for a Better Understanding of the Built Environment. M. Behnisch, G. Meinel (eds), *Trends in Spatial Analysis and Modelling*, Geotechnologies and the Environment, 19, Springer International Publishing, Cham, 167–191.

Wysocki, O., Schwab, B., Willenborg, B., 2022. Awesome CityGML, v1.0.0. https://doi.org/10.5281/zenodo.5899095.

Yakubovskiy, P., 2019. Segmentation models: Python library with neural networks for image segmentation based on keras and tensorflow. https://github.com/qubvel/segmentation_models.

Yu, J., Wang, Z., Majumdar, A., Rajagopal, R., 2018. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule*, 2(12), 2605–2617.

Zhang, W., Li, Z., Shan, J., 2021. Optimal Model Fitting for Building Reconstruction From Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9636-9650.