Object Detection for the Enrichment of Semantic 3D City Models with Roofing Materials

Lukas Arzoumanidis[1], Son H. Nguyen[1], Lara Johannsen[1], Filip Rothaut[1], Weilian Li[1,2], Youness Dehbi[1]

[1] Computational Methods Lab, HafenCity University, Hamburg, Germany - {firstname.lastname}@hcu-hamburg.de
[2] Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, China

Keywords: Object Detection, Roof Materials, Semantic Enrichment, 3D City Model, 3DCityDB, CityGML, OSM, Aerial Imagery

Abstract

Semantically rich 3D city models play a vital role in a variety of applications, such as urban planning. Enhancing these models with currently unavailable attributes, such as roof material types, can unlock new opportunities to tackle pressing challenges, including climate change mitigation and sustainable urban development. In this work, we present an end-to-end pipeline for the automatic detection of roof materials to semantically enrich 3D city models. To support this, a comprehensive training dataset was prepared by automatically annotating roof materials across Germany using OpenStreetMap (OSM) attributes and high-resolution orthophotos. Our object detection pipeline classifies five distinct roof material types using the YOLOv11-L architecture. Our detection results enabled the automatic augmentation of CityGML-based 3D models, filling in missing roof material information. This enrichment supports advanced applications, such as assessing roof suitability for green infrastructure or simulating urban heat island mitigation strategies. We validated the feasibility of our approach with real-world data and applied the method to a district in the city of Bremen, Germany. The paper also includes a detailed discussion of the learning process quality, the integration, and the visualization of the enriched 3D city model. The used code is available at: `https://github.com/hcu-cml/citydb-roofmats-ai`.

1. Introduction

The integration of cutting edge technologies and infrastructures is vital for enhancing urban sustainability. As urban populations continue to grow, cities must explore innovative approaches to meet the demands of their citizens while simultaneously reducing their carbon footprint for sustainable living environments. In this context, semantic knowledge of individual buildings across a wide range of attributes is essential for accurately estimating fossil fuel emissions and for enabling informed, data-driven planning aimed at their reduction (Sun et al., 2025; Park et al., 2024). This is particularly significant given that the building sector alone is estimated to account for approximately one-third of global carbon emissions (Sun et al., 2025).

Semantic information embedded in 3D building models, particularly rooftop characteristics, such as type or material, provide invaluable insights for urban climate analysis (Ilehag et al., 2018). This allows for the creation of controlled simulation environments, which are essential for evaluating the suitability of roofs for solar energy platforms, green infrastructure, and other climate-related interventions (Park et al., 2024). In this context, green rooftop infrastructure, such as Blue-Green Roof Systems or Phase-Change Materials (PCM) contribute to mitigating the Urban Heat Island (UHI) effect as they can increase evaporative cooling and store heat more effectively (Richter and Dickhaut, 2023; Roman et al., 2016). However, to simulate their impact effectively, urban planners rely on semantically enriched 3D city models, which could benefit significantly from the integration of semantic information, such as roof material composition (Nouvel et al., 2015). Furthermore, incorporating rooftop characteristics into semantic 3D city models enables urban planners to conduct more advanced simulations and analyses addressing a range of urban challenges not limited to climate-related issues, but also including emergency route planning, thereby supporting more informed and effective decision-making (Biljecki et al., 2015).

However, in practice, semantic information on roof materials is scarcely available and rarely included in publicly accessible semantic 3D city models. To address this need, our work proposes an approach for the automatic classification and subsequent semantic enrichment of roof material information within 3D city models using Volunteered Geographic Information (VGI). For this aim, we exploited the sparse information about roof materials from OpenStreetMap (OSM) in whole Germany to label a training dataset for the sake of object detection. The results open up new opportunities to augment and enrich city models with this a-priori unknown information for unmapped rooftops. Thus, we subsequently integrated the estimated roof materials, obtained from our learned model, into existing semantic 3D city models based on the City Geography Markup Language (CityGML) standard (Gröger et al., 2012; Kolbe et al., 2021). This end-to-end pipeline allows for an automatic city or even country-wide enrichment of 3D city models as baseline for further urban simulations of relevance. For instance, in the specific context of blue-green roof systems, the suitability of a building for roof greening heavily depends on the underlying material – concrete is typically more supportive due to its load-bearing capacity and thermal stability, whereas glass or metal roofs are often unsuitable without significant structural modification. Figure1 illustrates the resulting 3D city model highlighting different roof materials from our object detection in different colors accordingly.

The key contributions of this work are as follows: (1) **Roof Material Dataset:** Development of a roof material dataset derived from OSM data and high-resolution aerial imagery across Germany (2) **Object Detection:** Implementation of a city-wide roof material detection and mapping pipeline using a state-of-the-art deep learning-based object detection (3) **Semantic Enrichment:** Integration of the predicted roof material information into a semantic 3D city model in CityGML format. This integration is achieved by spatially aligning building footprints from the city model with their corresponding roof polygons derived from raster data, such as orthophotos.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W6-2025-9-2025

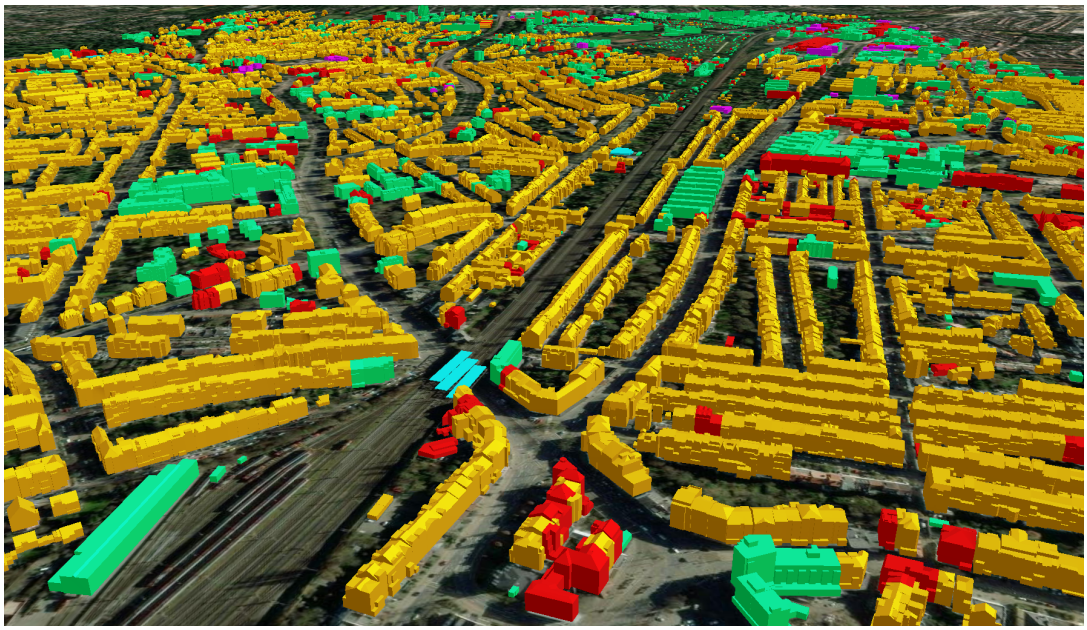


Figure 1. Visualization of buildings with enriched roof materials in the test area of Bremen, Germany. Color-coded 3D building models indicate their roof materials: concrete (grey), metal (magenta), glass (cyan), roof tiles (orange), tar paper (green), and multiple materials (red). Visualized using CesiumJS and 3DCityDB Web Map Client, with ArcGIS World Imagery as the basemap.

2. Automatic Extraction of Semantic Rooftop Knowledge

For many years, extensive research has been conducted on the automatic detection and segmentation of rooftops predominantly focusing on general rooftop classification and identification, e.g., using Support Vector Machines (SVMs) (Mohajeri et al., 2018) or Markov Random Fields (MRFs) (Katartzis and Sahli, 2008). Besides, rooftop reconstruction from 3D point clouds through prior knowledge and informed model sampling has also been explored (Dehbi et al., 2021). In this context, deep learning-based approaches proved state-of-the-art accuracy in many different semantic segmentation and classification tasks of various domains including rooftops. For instance, the application of Convolutional Neural Networks (CNNs) has significantly improved the automatic detection of rooftops and classification roof types from aerial images Alidoost and Arefi (2018). In a recent study, Yuan (2025) proposed an approach to address the challenges posed by buildings that often exhibit scale variation, spectral heterogeneity, and geometric complexity. The method integrates a multi-scale global perceptron network, combining Transformer and CNN architectures with modified encoder-decoder structures to enhance the contextual representation of buildings. In particular, the model employs an improved multi-head attention encoder to better capture spatial and spectral features.

To monitor health and integrity of flat rooftops, Abdullah et al. (2022) applied a CNN-based approach to detect faulty bolts, which indicate potential rooftop damage. Similarly, to infer the extent of rooftop damage, Hezaveh et al. (2017) harnessed a CNN to detect rooftop areas affected by hail impact. Moister stains on flat rooftops indicate ongoing degradation processes where water infiltration occurs, compromising the structural integrity of the roof and the durability of a building. Because manual inspections on large rooftops often can oversee such moister stains on flat rooftops, dos Santos et al. (2025) compared two deep learning-based approaches for the detection and segmentation of the aforementioned areas. Besides automatic approaches for rooftop segmentation and classification, methods for the automatic detection of rooftop infrastructure, such as photovoltaic panels (Wang et al., 2023) and the subsequent potential analysis (Wu and Biljecki, 2021) from aerial imagery, have been researched plentiful, proposing deep learning-based approaches to achieve state-of-the-art accuracies.

In the specific context of roofing materials, Ilehag et al. (2018) compare various segmentation approaches for predicting roof materials using high-resolution multispectral and thermal infrared imagery. Their work distinguishes between two material classes: cement tiles and a combined class of colorbond and zincalume. In a subsequent step, the predicted roof materials are integrated into a 3D model of a specific council area in Perth, Australia. Similarly, Solovyev (2020); Kim et al. (2021) employed a deep learning-based segmentation to classify roof materials in a small areas of interest. Their model was trained to identify five classes: concrete and cement, healthy metal, irregular metal, and incomplete rooftops. Aiming to detect asbestos–cement roofing in aerial imagery, Krówczyńska et al. (2020) developed a Convolutional Neural Network (CNN)-based approach using both RGB and color-infrared image compositions featuring a spatial resolution of 25 cm. In their study, Park et al. (2024) utilize aerial imagery to estimate the potential energy savings and construction costs associated with implementing cool roof systems in large urban environments. Their approach employs a Convolutional Neural Network (CNN) to detect building rooftops, which are then clustered based on surface color, serving as a proxy for estimating the energy-saving potential of cool roof retrofitting. Similarly, Sun et al. (2025) developed a deep learning-based method to automatically identify roof and facade materials using remote sensing data combined with Google Street View imagery.

Biljecki and Dehbi (2019) used available roof attributes from OSM to label and predict roof types. Similarly, we annotated a dataset using roof material properties from OSM for end-to-end detection of different material classes from aerial imagery. However, we used a deep learning-based approach, i.e., YOLOv11, instead of manually designing relevant features.

class	train dist.	val dist.	test dist.	total obj.
tar paper	17,843	4069	4011	25,923
concrete	1715	395	438	2548
metal	2853	668	652	4173
glass	2055	404	482	2941
roof tiles	107,335	23,022	23,277	153,634
total obj.	131,801	28,558	28,860	189,219

Table 1. Targeted classes and their respective distribution in the dataset.

3. Methodology

The following section presents our proposed pipeline for detecting, mapping, and integrating roof materials into CityGML.

3.1 Multi-Class Object Detection

After extracting building roof material information from OSM, we convert it along with the corresponding aerial imagery into the YOLO object detection format. The resulting dataset is then randomly divided into three subsets: $80\,\%$ training, $10\,\%$ validation, and $10\,\%$ testing. Using the training subset, we train the deep learning-based YOLOv11-L model to simultaneously detect five distinct roof material classes, as illustrated in Figure 2. The five roof material classes targeted in our study are tar paper, concrete, metal, glass, and roof tiles totaling $189,219$ instances in our dataset. Further details on their distribution among the dataset are provided in Table 1.

This classification enables the model to distinguish between these materials during detection and allows for the individual evaluation of each class's performance. The classes were selected based on their prominence in Germany. For our experiments, we employed the YOLOv11 implementation developed by Ultralytics[1]. Specifically, we selected the *Large* version of YOLOv11, which had been pretrained on the COCO dataset prior to our training and fine-tuning and more learnable parameters compared to the *Regular* and *Small* versions, respectively. The model was chosen for its broad applicability and widespread use. Compared to earlier versions, YOLOv11 enhances spatial attention, allowing the model to concentrate on key regions of an image. This improvement helps achieve more accurate detections, particularly for smaller or partially occluded objects (Khanam and Hussain, 2024).

During training, we monitor performance using three loss components: bounding box loss (box loss), classification loss (cls loss), and Distributed Focal Loss (DFL loss). Classification loss quantifies the discrepancy between the true class label and the class predicted by the model. Bounding box loss evaluates the alignment between predicted and ground-truth bounding boxes, incorporating factors such as aspect ratio and the spatial distance between box centers. This is typically achieved using an IoU-based loss function. In addition, DFL is incorporated to enhance the learning of bounding box coordinates by reformulating the task as a classification problem rather than a direct regression. Instead of predicting a continuous value for each coordinate, the model outputs a discrete probability distribution over a predefined range of n values where $\hat{y}$ corresponds to the final prediction of the coordinates of a bounding box and y_i represents different candidate values for a coordinate as can be seen in Equation 1. Focal loss is then applied to this distribution to prioritize high-confidence predictions while down-weighting less informative ones, thereby improving localization accuracy (Li et al., 2020).

$$\hat{y} = \sum_{i=0}^{n} P(y_i) y_i \tag{1}$$

For post-training evaluation, we employ Precision, Recall, and mean Average Precision (mAP) as metrics. As of the time of writing, mAP is a widely used metric for assessing model performance in object detection and information retrieval tasks. In our study, we utilize mAP to evaluate the accuracy of predicted bounding boxes in comparison to the annotated ground-truth data. The following equation describes the mAP formally:

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^{k=n} \text{AP}_i, \tag{2}$$

where n corresponds to the number of classes and AP_i is the average precision of class i accordingly. To compute mAP, it is necessary to first calculate the Intersection over Union (IoU), along with the Precision and Recall metrics. IoU measures the ratio between the area of overlap and the area of union between the predicted bounding box and the ground-truth bounding box. Precision is defined as the ratio of true positive detections to the total number of predictions made, while Recall is the ratio of true positive detections to the total number of actual positives. To assess the quality of predictions, an IoU threshold is applied beforehand so that only predictions with sufficient overlap with the ground-truth are considered true positives. The AP_i in Equation 2 corresponds to the area under the Precision-Recall (PR) curve for each class. This curve is generated by ranking model predictions by their confidence scores and computing Precision and Recall at each threshold level. Finally, the AP values are averaged across all classes to obtain the mAP. In some cases, mAP is further evaluated across multiple IoU thresholds; for instance, mAP_{50} refers to the average precision computed at a fixed IoU threshold of 0.5.

3.2 Roof Material Integration in Semantic 3D City Models

The City Geography Markup Language (CityGML) is a standardized information model and data exchange format issued by the Open Geospatial Consortium (OGC). It is designed to model, store, and exchange detailed representations of a broad range of 3D urban and landscape features (Gröger et al., 2012). Unlike other 3D city models that focus primarily on visual aspects, such as those based on aerial mesh geometries, CityGML integrates appearance, geometry, and topology with rich semantic information. The latest version, CityGML 3.0, was released as an international OGC standard between 2021 and 2023 (Kolbe et al., 2021; Kutzner et al., 2023).

In practice, many applications do not operate directly on text-based CityGML datasets. Instead, these datasets are typically parsed and transformed into structured representations that can be efficiently stored, managed, and queried using a database management system. Such representations may take the form of object-oriented, graph-based, or relational data models.

Relational models represent information as collections of tables (relations) composed of rows (tuples) and columns (attributes). Since the advent of Structured Query Language (SQL), Relational Database Management Systems (RDBMSs) have become standard across many domains (Davoudian et al., 2018). This popularity has led to extensive research on storing and managing XML-based formats like GML and CityGML in relational systems. Several RDBMSs, including Oracle, IBM Db2,

[1] `https://github.com/ultralytics/ultralytics?tab=readme-ov-file`

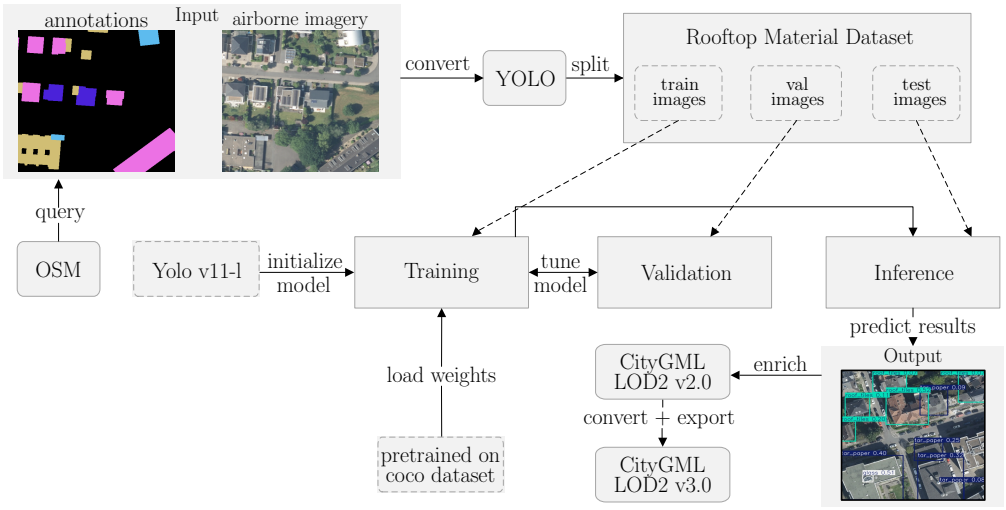


Figure 2. Overview of our approach for detecting, integrating and mapping building roof materials into semantic 3D city models.

PostgreSQL, and Microsoft SQL Server, support native XML or GML handling. Spatial extensions, like PostGIS for PostgreSQL, further enable processing of geographic data. CityGML can be managed using relational models. For example, the database *cjdb* (Powałka et al., 2024) leverages CityJSON (Ledoux et al., 2019) for compact storage in PostgreSQL. The most widely used software, the 3D City Database (3DCityDB) (Yao et al., 2018), is an open-source, high-performance geo-database for managing, visualizing, and exporting large-scale CityGML datasets.

Since May 2025, the 3DCityDB has been available in its latest major release, version 5.0, which offers full support for CityGML 3.0 and introduces a fully redesigned and simplified database schema, resulting in improved performance and greater flexibility. With the 3DCityDB's well-established software suite and the promising capabilities of this new version, we chose to adopt version 5.0 and its associated tools to store and manage the selected CityGML test dataset for Bremen. At the same time, this study serves as one of the first to explore and evaluate the new release in practice.

The redesigned database schema of the 3DCityDB now includes a single table *feature* for all feature types (like buildings) and a single table *property* for most semantic attributes and associations (see Figure 3). The CityGML data model supports the storage of arbitrary attribute values with user-defined names through the concept of generic attributes. This mechanism allows for storing the predicted roof materials for each building directly within the database without requiring prior schema extensions or additional tables.

The integration of predicted roof materials into the existing semantic 3D city model is performed in two steps: (1) roof segments identified in the orthophotos are matched to corresponding buildings in the city model based on spatial overlap, such as by comparing their bounding boxes, and (2) the predicted roof materials are then inserted as generic string attributes into the associated building objects within the database. The enriched city model can then be exported to both CityGML 2.0 and 3.0.

To accurately associate an extracted roof segment, represented by a bounding box A, with a corresponding building from the city model, bounded by bounding box B, we employ the Jac-

feature

id	objectid	envelope	...
1	DEHB01ALn0001O0e	01030000_	...
25	DEHB01ALn0001PBB	01030000_	...
43	DEHB01ALn0001O2t	01030000_	...

Primary key

property Foreign key

id	feature_id	name	val_string	...
7	1	DatenquelleLage	1000	...
10	1	status	measured	...
14	1	roofType	2100	...

Figure 3. Example of the contents and the primary key - foreign key relationship between the tables *feature* and *property*.

card index (also known as IoU), defined as the ratio of the intersection area to the union area of the two geometries:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{3}$$

A match is considered valid if the Jaccard index exceeds a predefined threshold. However, as illustrated in Figure 4, mismatches can arise in two directions: (1) a single bounding box in the orthophoto may cover multiple buildings in the city model, and (2) a single building in the city model may intersect with several smaller roof bounding boxes from the orthophoto. In such cases, the standard Jaccard index may fail to establish desired results. To address this, we shift the matching strategy to prioritize building bounding boxes, thereby maximizing the number of correctly assigned roof materials in the city model. Specifically, we employ a modified version of the Jaccard index in which the denominator is limited to the area of the building bounding box only:

$$J'(A, B) = \frac{|A \cap B|}{|A|}. \tag{4}$$

This approach accommodates cases where the building bounding box is fully contained within a larger roof segment, as well as instances where it overlaps partially with multiple smaller roof segments. The threshold for $J'(A, B)$ determines the sens-

Figure 4. Example cases where a segmented roof may contain multiple buildings (left), and vice versa, where a building may contain multiple detected roofs (right).

itivity of the matching: lower values increase the number of matches (at the risk of false positives), while higher values improve precision but reduce potential matches. In this study, a threshold of $J'(A, B) = 0.15$ is used. Both the standard and modified forms of the Jaccard index, including the variant used in this study, have been widely employed in related spatial matching applications (Ding et al., 2024). In PostgreSQL/PostGIS, spatial overlapping is done using an R-tree index.

After a segmented roof and its corresponding building have been successfully matched, the predicted roof material is stored as a generic attribute of the building using the following SQL statement:

```
INSERT INTO citydb.property (
   feature_id, datatype_id, namespace_id,
   name, val_string)
VALUES (id, 5, 3, 'PredictedRoofMaterials',
   'concrete|metal|glass|roof_tiles|tar_paper');
```

The table *property* is defined in the schema *citydb*, while the values *datatype_id* and *namespace_id* are 5 and 3 for a generic string attribute in the 3DCityDB version 5.0, respectively.

4. Experimental Results

The following section describes the generation of training data in more detail and gives insight into the experimental results our approach achieved evaluating commonly used quantitative metrics for object detection.

4.1 Annotated Training Data Generation

To generate annotated training data containing roof material information, we utilized OSM. To ensure coverage of all commonly found materials in Germany, we queried the complete OSM dataset for the country using the Overpass API. In the next step, we constructed a grid with cells of 100×100 meters, which served to define both the spatial extent of the roof material annotations and the corresponding aerial image patches for training. To focus on the most relevant data, we filtered the OSM-derived material types to retain only the most frequently occurring ones, removing rare or ambiguous entries. Additionally, we imposed a constraint on grid cell selection: only those cells containing at least three distinct rooftop instances, each larger than $10\,\mathrm{m}^2$, were retained. This filtering step aimed to reduce dataset size while ensuring that training samples contained dense and meaningful information.

Finally, we used the selected grid cells to extract aerial image patches at a map scale of 1:2400. Each patch was cropped to a size of 500×500 pixels, resulting in the final training pairs of aligned image data and roof material labels. All aerial imagery was openly provided by the geoinformation authorities of the federal states of Germany and features a ground sampling distance (GSD) of $20\,\mathrm{cm}$. Since many roofs are only a few meters in size, high-resolution RGB aerial imagery is preferred over lower-resolution hyperspectral data. This is based on the assumption that the color and patterns of roof materials in high-resolution imagery are more informative for our model compared to the less detailed features in hyperspectral data.

To visualize the predicted results in our area of interest, we converted the object coordinates from the YOLO annotation format to real-world coordinates in the ETRS89 / UTM Zone 32N coordinate system. This conversion was performed using the coordinate bounding box provided with each aerial image, which is stored as a GeoTIFF. The resulting coordinates were then saved as Well Known Text (WKT) for the integration into 3D city models, such as CityGML, later on.

4.2 Experimental Setup

To detect, classify and map different classes of roof material we apply as mentioned a pretrained deep learning-based object detection model, namely, YOLOv11-L with 25.3 million parameters and 86.9 billion Floating-Point Operations Per Second (FLOPS) provided by Ultralytics. The model has been pretrained on the COCO dataset[2], featuring 80 object classes. The model is implemented in PyTorch where the CUDA accelerator has been used. In this context, our experiments have been conducted with PyTorch version 2.2.0 with CUDA version 12.8 using an NVIDIA RTX 2000 Ada GPU with 8GB of VRAM. The default configuration of hyperparameters have been select and the model was trained for 100 epochs.

For the integration of predicted roof materials, we selected the publicly available[3] CityGML 2.0 dataset of Bremen at Level of Detail (LoD) 2. Specifically, we focused on a $4\,\mathrm{km}$ by $3\,\mathrm{km}$ area encompassing the districts of Mitte, Östliche Vorstadt, Schwachhausen, and Vahr.

4.3 Multi-Class Object Detection

The experimental results reveal distinct performance variations across training, validation, and testing phases for all five target classes. During training and validation, all three loss functions exhibit a consistent downward trend over the course of the epochs. As shown in Figure 6a & 6b, the bounding box loss and classification loss follow a similar range of values, in contrast to the DFL loss. This similarity is attributed to the fact that both losses evaluate the model's performance on regression tasks locating objects and bounding boxes in the case of the bounding box loss, and predicting the correct class labels in the case of the classification loss, as described in Section 3.1. When comparing the training and validation bounding box loss with the classification loss further, it becomes evident that after 50 epochs, the classification loss decreases more rapidly than the bounding box loss. When evaluating the mAP, mAP_{50}, Precision, and Recall during validation, all metrics demonstrate an increasing trend across epochs, as illustrated in Figure 6c. The consistent decrease in both training and validation losses,

[2] https://cocodataset.org
[3] https://metaver.de

coupled with the increase in validation metrics, indicates effective model learning and generalization.

To evaluate our trained model on the test dataset, we computed the F1-Confidence curve, Precision-Recall curve, and confusion matrix, which collectively illustrate the model's performance across all five target classes, as shown in Figure 7. The F1-Confidence curve (Figure 7a) captures the model's accuracy by combining Precision and Recall scores across varying confidence thresholds. Our results show that the highest overall mean F1 score across all classes is 0.62, achieved at a confidence threshold of 0.286. This is a reasonably good result, considering the heterogeneity of the dataset, as discussed in Section 4.1. A closer examination of individual classes reveals that while metal, glass, and tar paper exhibit similar trends in their F1 scores across confidence thresholds, roof tiles and concrete emerge as outliers.

As shown in Table 1, class imbalance has a notable impact on model performance. The roof tiles class exhibits significantly higher F1 scores, exceeding the overall mean F1 score across all classes by approximately 20 %. In contrast, the concrete class which contains substantially fewer samples shows markedly lower F1 scores, falling roughly 12 % below the overall mean. Furthermore, the F1-confidence curve reveals that the highest F1 score for roof tiles is achieved at a comparatively higher confidence threshold. This indicates that the model can achieve more accurate and confident predictions for the roof tiles class relative to the average performance across all classes.

To further support the previous observation, Figure 7b presents the Precision-Recall curve. Consistent with the F1-Confidence curve, the Precision-Recall curve shows notably higher Precision and Recall values for the roof tiles class, while the concrete class exhibits lower performance. Additionally, as observed earlier, the metal, glass, and tar paper classes perform close to the overall class mean. The highest mAP_{50} score representing the area under the Precision-Recall curve at an IoU threshold of 0.50 is 0.624, indicating strong overall performance across all confidence thresholds and for all classes combined.

As shown in Table 2, the mAP for both the metal and glass classes is 0.604, which may suggest a weak correlation with the number of training samples. Both classes had a similar number of samples extracted from OSM, as indicated in Table 1. This potential relationship is further supported by the roof tiles class, which achieved the highest mAP value of 0.841 and also had the largest number of training samples. However, the concrete class, despite having a comparable number of training samples to metal and glass, shows notably lower performance. This suggests that factors beyond sample size may influence the model's performance. One possible explanation is the greater spectral heterogeneity of concrete rooftops in RGB imagery. Concrete roofs are often covered with additional materials such as tar or gravel to enhance weather resistance, resulting in a more variable visual appearance compared to glass or metal surfaces.

Interestingly, across all classes, the most frequent misclassification occurs with the background class, as shown in Figure 7c. This observation is important, as it suggests that the primary source of misdetections is not confusion between material classes, but rather a failure to detect any object at all, classifying them as background. This may imply that class imbalance in the dataset is not the primary factor limiting model performance, since errors are predominantly due to missed detections rather than incorrect class assignments among roof material categories.

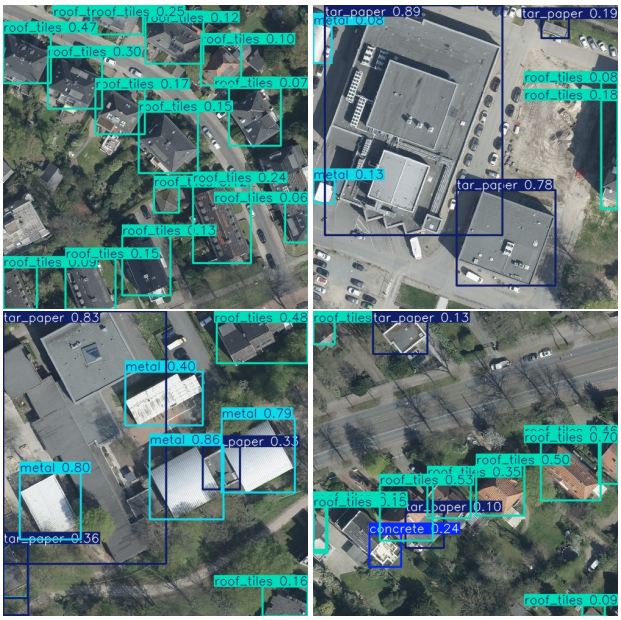


Figure 5. Examples of predicted roof material objects identified in four aerial images.

As illustrated in Figure 5, the model successfully detects nearly all roof material types across all classes in each example image, with the exception of rooftops that are partially occluded by vegetation. A likely explanation for this limitation is that the aerial imagery used is based solely on high-resolution RGB data rather than hyperspectral imaging. As a result, the model lacks the spectral depth necessary to infer or distinguish materials that are obscured by vegetation, making it difficult to identify what lies beneath these occluded areas.

4.4 Integration in CityGML

Out of 20,764 buildings contained in the selected CityGML dataset of Bremen, 17,139 buildings have been assigned with at least a roof material, as illustrated in Figure 1. This corresponds to a coverage of 83 %. Among these, approximately 78 % have been assigned with the roof material roof tiles, while 16 % are labeled with tar paper. Buildings with multiple roof materials account for 5 %, and the remaining 1 % consist of buildings with a single roof material, such as concrete, metal, or glass. A breakdown of these distributions is presented in Figure 8. In total, 17,994 generic string attributes representing roof materials have been inserted into the city model, with 828 buildings assigned more than one material.

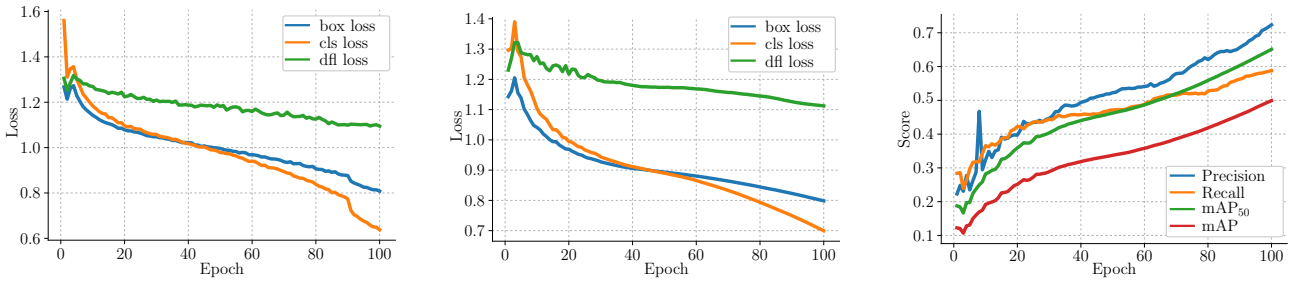
5. Conclusion & Outlook

In this paper, we addressed the automatic enrichment of semantic 3D city models with roof material information. Using

class	Precision	Recall	mAP	mAP_{50}
tar paper	0.662	0.58	0.632	0.49
concrete	0.618	0.377	0.439	0.336
metal	0.725	0.495	0.604	0.455
glass	0.719	0.544	0.604	0.432
roof tiles	0.742	0.837	0.841	0.664
all	0.693	0.567	0.624	0.475

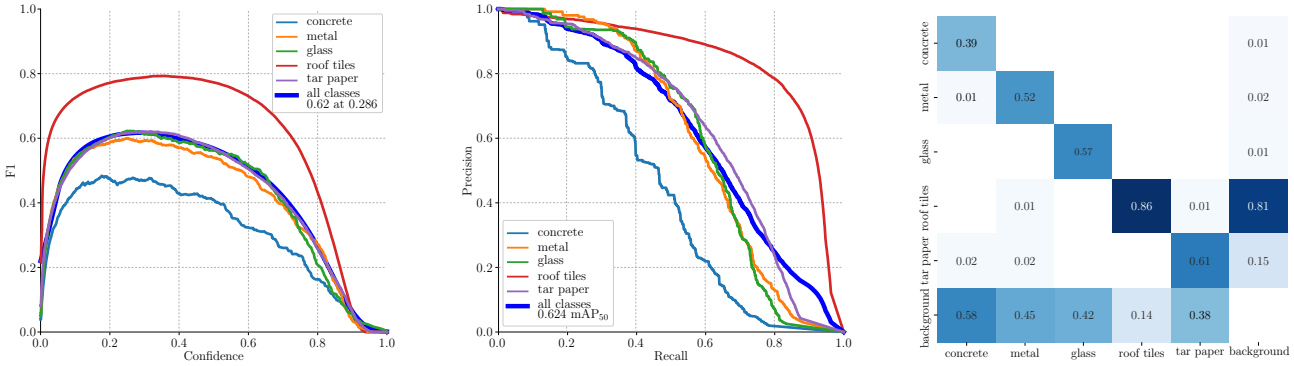
Table 2. Classes and their respective bounding box Precision, Recal, mAP and mAP_{50} after inference on the test dataset.

https://doi.org/10.5194/isprs-annals-X-4-W6-2025-9-2025



(a) training loss (b) validation loss (c) validation metrics

Figure 6. Training and validation loss, validation mAP, Precision and Recall.



(a) F1-Confidence (b) Precision-Recall (c) normalized confusion matrix

Figure 7. F1-Confidence, Precision-Recall and confusion matrix of targeted classes after inference on the test dataset.

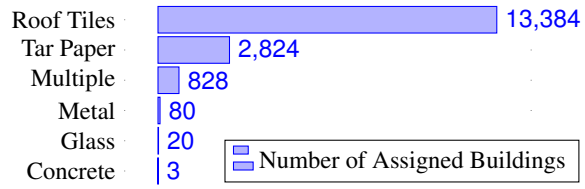


Figure 8. Distribution of buildings by assigned roof materials.

volunteered geographic information (VGI) from OpenStreetMap (OSM) across Germany, where roof material attributes are available, and high-resolution orthophotos, we created a labeled dataset for object detection. This enabled the categorization of roof materials, facilitating the enrichment of existing 3D city models. As a result, our approach allows for the large-scale expansion of CityGML datasets, supporting applications like roof greening initiatives and urban heat island assessments. Our approach is applicable across Germany and can be extended to regions beyond, provided they exhibit similar building morphologies. Alternatively, the model can be adapted for global application, contingent upon the availability of high-resolution orthophotos and roof materials provided through VGI.

To achieve even more robust detection, the class imbalance problem in the underlying OSM attributes could be addressed as a next step. In particular, integrating roof type information into the learning process could enhance the prediction quality for specific classes. The investigation of the impact of incorporating high-resolution hyperspectral images will be also a subject of future research. The achieved material classification could also enable new applications beyond roofs, such as the estimation or prediction of the underlying building materials and thermal insulation. For the sake of bidirectional data enrichment, we can also augment the roof material attributes from the CityGML model into OSM.

The compact schema of the new 3DCityDB simplifies queries, enabling the insertion of roof materials for 20,000 buildings in Bremen within minutes. Due to inherent schema limitations, these attributes were added as generic atrtibutes. In the near future, we also aim to explore graph-based approaches, supported by recent advancements in CityGML graphs.

Acknowledgements

This paper was supported by the Next Generation City Networks project (Grant No 19DZ24004C) founded by the Federal Ministry for Digital and Transport.

References

Abdullah, M., Ahmad, M., Kim, Y.-G., Han, D., 2022. Small Objects and Faults Detection on Corrugated Metal Roof using Drone equipped with Deep Learning. *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 1–4.

Alidoost, F., Arefi, H., 2018. A CNN-Based Approach for Automatic Building Detection and Recognition of Roof Types Using a Single Aerial Image. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86(5), 235-248.

Biljecki, F., Dehbi, Y., 2019. Raise the Roof: Towards Generating LOD2 Models without Aerial Surveys using Machine Learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W8, 27–34.

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.

Davoudian, A., Chen, L., Liu, M., 2018. *A Survey on NoSQL Stores*. 51Number 2, Association for Computing Machinery (ACM), 1–43.

Dehbi, Y., Henn, A., Gröger, G., Stroh, V., Plümer, L., 2021. Robust and Fast Reconstruction of Complex Roofs with Active Sampling from 3D Point Clouds. *Transactions in GIS*, 25(1), 112–133.

Ding, L., Xiao, G., Pano, A., Fumagalli, M., Chen, D., Feng, Y., Calvanese, D., Fan, H., Meng, L., 2024. *Integrating 3D City Data through Knowledge Graphs*. Informa UK Limited, 1–20.

dos Santos, L. M. A., Lescano, L. R., Higa, G. T. H., Zanoni, V. A. G., da Silva, L. S., Alvarez, C. I., Pistori, H., 2025. Mapping Stains on Flat Roofs using Semantic Segmentation based on Deep Learning. *Case Studies in Construction Materials*, 22.

Gröger, G., Kolbe, T. H., Nagel, C., Häfele, K.-H., 2012. *OGC City Geography Markup Language (CityGML) Encoding Standard*. Open Geospatial Consortium.

Hezaveh, M. M., Kanan, C., Salvaggio, C., 2017. Roof Damage Assessment using Deep Learning. *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 6403–6408.

Ilehag, R., Bulatov, D., Helmholz, P., Belton, D., 2018. Classification and Representation of Commonly used Roofing Material Using Multisensorial Aerial Data. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1, 217-224.

Katartzis, A., Sahli, H., 2008. A Stochastic Framework for the Identification of Building Rooftops using a Single Remote Sensing Image. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1), 259-271.

Khanam, R., Hussain, M., 2024. YOLOv11: An Overview of the Key Architectural Enhancements.

Kim, J., Bae, H., Kang, H., Lee, S. G., 2021. CNN Algorithm for Roof Detection and Material Classification in Satellite Images. *Electronics*, 10(13).

Kolbe, T. H., Kutzner, T., Smyth, C. S., Nagel, C., Roensdorf, C., Heazel, C., 2021. *OGC City Geography Markup Language (CityGML) Version 3.0 Part 1: Conceptual Model Standard*. Open Geospatial Consortium. International Standard.

Krówczyńska, M., Raczko, E., Staniszewska, N., Wilk, E., 2020. Asbestos-Cement Roofing Identification Using Remote Sensing and Convolutional Neural Networks (CNNs). *Remote Sensing*, 12(3).

Kutzner, T., Smyth, C., Nagel, C., Coors, V., Vinasco-Alvarez, D., Ishimaru, N., Yao, Z., Heazel, C., Kolbe, T. H., 2023. *OGC City Geography Markup Language (CityGML) Version 3.0 Part 2: GML Encoding Standard*. Open Geospatial Consortium. International Standard.

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., Vitalis, S., 2019. *CityJSON: A Compact and Easy-to-use Encoding of the CityGML Data Model*. 4Number 1, Springer Science and Business Media LLC.

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., Yang, J., 2020. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Curran Associates Inc., Red Hook, NY, USA.

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., Scartezzini, J.-L., 2018. A City-scale Roof Shape Classification using Machine Learning for Solar Energy Applications. *Renewable Energy*, 121, 81-93.

Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V., Eicker, U., 2015. Combining GIS-based Statistical and Engineering Urban Heat Consumption Models: Towards a new Framework for Multi-scale Policy Support. *Energy and Buildings*, 107, 204-212.

Park, J., Park, S., Kang, J., 2024. Detecting and Classifying Rooftops with a CNN-based Remote-sensing Method for Urban Area Cool Roof Application. *Energy Reports*, 11, 2516-2525.

Powałka, L., Poon, C., Xia, Y., Meines, S., Yan, L., Cai, Y., Stavropoulou, G., Dukai, B., Ledoux, H., 2024. *cjdb: A Simple, Fast, and Lean Database Solution for the CityGML Data Model*. Lecture Notes in Geoinformation and Cartography, Springer Nature Switzerland, Technical University of Munich (TUM), Munich, Germany, 781–796.

Richter, M., Dickhaut, W., 2023. Long-Term Performance of Blue-Green Roof Systems—Results of a Building-Scale Monitoring Study in Hamburg, Germany. *Water*, 15(15).

Roman, K. K., O'Brien, T., Alvey, J. B., Woo, O., 2016. Simulating the Effects of Cool Roof and PCM (Phase Change Materials) Based Roof to Mitigate UHI (Urban Heat Island) in Prominent US Cities. *Energy*, 96, 103-117.

Solovyev, R. A., 2020. Roof Material Classification from Aerial Imagery. *Optical Memory and Neural Networks*, 29(3), 198-208.

Sun, K., Li, Q., Liu, Q., Song, J., Dai, M., Qian, X., Gummidi, S. R. B., Yu, B., Creutzig, F., Liu, G., 2025. Urban Fabric Decoded: High-precision Building Material Identification via Deep Learning and Remote Sensing. *Environmental Science and Ecotechnology*, 24, 100538.

Wang, J., Chen, X., Shi, W., Jiang, W., Zhang, X., Hua, L., Liu, J., Sui, H., 2023. Rooftop PV Segmenter: A Size-aware Network for Segmenting Rooftop Photovoltaic Systems from High-resolution Imagery. *Remote Sensing*, 15(21).

Wu, A., Biljecki, F., 2021. Roofpedia: Automatic Mapping of Green and Solar Roofs for an Open Roofscape Registry and Evaluation of Urban Sustainability. *Landscape and Urban Planning*, 214, 104167.

Yao, Z., Nagel, C., Kunde, F., Hudra, G., Willkomm, P., Donaubauer, A., Adolphi, T., Kolbe, T. H., 2018. 3DCityDB - A 3D Geodatabase Solution for the Management, Analysis, and Visualization of Semantic 3D City Models based on CityGML. *Open Geospatial Data, Software and Standards*, 3(5), 1-26.

Yuan, Q., 2025. Building Rooftop Extraction from High Resolution Aerial Images using Multiscale Global Perceptron with Spatial Context Refinement. *Scientific Reports*, 15(1), 6499.