A Concept for Methodical Classification, Assignment of Attributes, and Labeling of Facade Elements in Camera Images for LOD3 Reconstruction of Buildings

FLORIAN FRANK[1], RICO RICHTER[2] & LUDWIG HOEGNER[3]

Abstract: Urban challenges like climate change, population growth, and mobility demand advanced planning using 3D city models. Accurate simulations require detailed 3D models, posing cost challenges for LOD3 façade descriptions. This paper proposes a classification approach, annotation guideline, and examples for high-fidelity façade segmentation in 2D camera RGB images, aiding photogrammetric building reconstruction. Previously captured high-resolution panorama images with precise global 6DoF pose and position were cross-compared to CityGML and IFC, resulting in a hierarchically inherited classification. A detailed annotation guideline ensures a comprehensive dataset. Examples showcase the complexity of annotating façade components. With this proposal, the paper contributes to filling the gap between images and detailed façade component reconstruction to standardized outputs.

1 Introduction

Cities and urban areas are subject to permanent change and have an inherent need to address future challenges such as climate change, growing populations, and mobility. 3D city models are well established and used for simulation and planning tasks to enhance the overall quality of life. Typical applications include building energy loss analysis, renovation planning, emergency access for fire workers, wind simulation, heat spot simulation, automated driving simulation (SCHREINER et al. 2021), and autonomous driving map generation (TANG et al. 2022). All these applications could benefit from detailed facade information.
Accurate and predictive simulations require detailed 3D city models with a high level of detail, such as CityGML LOD3 building models (GRÖGER et al. 2021). Today, most German buildings are available as CityGML LOD2, representing buildings as so-called folded cardboard boxes. These building representations can be derived easily from nationwide aerial remote sensing data. Further, CityGML LOD3 and IFC (ISO 16739-1 2018) describe the outer appearance of buildings in semantics, geometry, and appearance. These representations can be assumed to be the most detailed level and output goal for a complete façade description. Available LOD3 real-world city models are rare due to expensive digitalization sensor setups and data acquisition cost impacts (WYSOCKI et al. 2022; SCHWAB et al. 2021).
Extending the state of digitalization of buildings in LOD3 challenges this cost-benefit problem, by making data readily available. Therefore, the broad application of different technologies, image-based data acquisition with cameras, and practical data is required. Another factor in the

[1] Institute Economy and Technology IWT, Mobility Department, Fallenbrunnen 14, D-88045 Friedrichshafen, E-Mail: Frank@iwt-bodensee.de
[2] Point Cloud Technology GmbH, August-Bebel-Str. 26-53, D-14482 Potsdam, E-Mail: Rico.Richter@pointcloudtechnology.com
[3] Hochschule München, Laboratory for Photogrammetry and Remote Sensing, Karlstr. 6, D-80333 München, E-Mail: Ludwig.Hoegner@hm.de

success of reconstructed buildings is their usability. The output data formats should be standardized and oriented to CityGML and IFC.

Research related to raw data façade component segmentation regarding standardized data formats was done in façade datasets and façade instance segmentation. However, no appropriate and most parts covering classification and dataset for image detailed façade segmentation and reconstruction were available.

Based on these circumstances, high-fidelity 2D RGB images of Campus Fallenbrunnen in Friedrichshafen with precise 6DoF positions and poses were taken. The visible details in the images are higher than the referenced output standards and allow the annotation of more functional details and attributes, like exact glass surfaces, door opening directions, or doorbells. To bridge the gap from acquired image data to standardized output, this paper proposes a high-fidelity classification and attribute assignment for façade component instance segmentation in 2D RGB images for future photogrammetric detailed outer building reconstruction in LOD3 quality.

2 Related Work

The raw 2D RGB image data of façades and idealized building façade representation like CityGML LOD3 and IFC relates to works of façade instance segmentation datasets and algorithms. Table 1 compares multiple published data sets with façade instance segmentation. Table 1 is an extended excerpt of WANG et al. (2023) CMP compared to other image-based instance façade segmentation datasets. The comparison is extended by SynthCity (GRIFFITHS & BOEHM 2019) as an artificial dataset, the TUM Façade mobile laser scan (MLS) dataset (WYSOCKI et al. 2022), and Solo V2 Façade (LU et al. 2023).

Tab. 1: Comparison of instance façade segmentation datasets. 'x' as yes; '~' as partially; '-' as not.

		Classification					
Dataset	**Data**	**Façade-segmentation detail-level**	**Inheritance**	**Classes**	**Standard oriented**	**Reconstruction oriented**	**Labels**
ECP 2011 (Teboul et al., 2010)	img.	~	1	7	-	-	Walls, windows, doors, balconies, roofs, stores, sky
Graz (Riemenschneider et al., 2012)	img.	~	1	4	-	-	Door, window, wall, sky
CMP (Tyleček and Šára, 2013)	img.	~	1	12	-	-	Façade, molding, cornice, pillar, window, door, sill, blind, balcony, shop, deco, background
ENPC 2014 (Lotte et al., 2018)	img.	~	1	7	-	-	Same as ECP 2011
RueMonge (Riemenschneider et al., 2014)	img.	~	1	7	-	-	Door, shop, balcony, window, wall, roof, and sky.
eTRIMs (Korč and Förstner, 2009)	img.	~	1	8	-	~	Building, car, door, pavement, road, sky, vegetation, window
LabelMeFacade (Frohlich et al., 2010)	img.	~	1	8	-	-	Building, car, door, pavement, road, sky, vegetation, window
CFP 2023 (Wang et al., 2023)	img.	~	1	9	-	x	Building, window, door, roof, tree, sky, people, car, and sign
SynthCity (Griffiths and Boehm, 2019b)	img.	~	1	7	-	-	Building, road, pavement, ground, natural ground, tree, pole-like, street furniture, Car
TUM Façade (Wysocki et al., 2022)	pts.	~	1	17	x	N/A	Wall surface, door, balcony, CityGML Building Installation
Solo V2 (Lu et al., 2023)	img.	~	1	2	-	x	Wall, window
Ours (LevelUp)	img.	x	3	39	x	x	Environment (Living Beings, Vegetation,...), Building (Boundaries, Additions,...) Infrastructure (Roads, Signals, Signs)

In comparison, many datasets consider façade object classes from the point of usage and distinguish objects in doors, windows, balconies, and shops. Our proposed object-class detail level goes beyond and distinguishes façades objects on a functional level. For example, windows consist of a fixed frame, a moveable casement, and a glass insert. They result in an

inheritance level of three and 39 classes. These proposed classes as well as the TUM Façade (WYSOCKI et al. 2022) classes are standard format oriented to CityGML and IFC in order to improve usage. Considering standardized formats, a preparation for high-level facade reconstructions is aimed for. The CFP dataset is an off-topic of façade parsing (WANG et al. 2023), and Solo V2 (LU et al. 2023) did not publish the dataset. Both methods aim to reconstruct the façade on the image level. However, the image-based datasets do not have a high detail level for façade instance segmented-based 3D reconstruction.
The related input to derive the high detail level of classification is based on high resulted 360° panorama images with exact global six-degree freedom pose (FRANK et al. 2023).

3 Methodology

The methodology section is splitting in the preliminary work of raw data input and the description of class definition for façade component instance segmentation including classes and their graph with additional attributes.

3.1 Raw Data Input

High-detail level output requires high-resolution and precise input data. This input data is delivered by the gimbal-based camera system on top of a handcart (FRANK et al. 2023). The original idea of the handcart lies in façade and building reconstruction in CityGML LOD3 quality. The raw output data of the handcart are mono camera images, GNSS+RTK, gimbal position, and inertial measuring unit (IMU) data. This data allows us to create 360° panorama images with a pixel and poses an angular resolution of 20 arcseconds. The global position is accurate in the millimetre range. These panorama images allow us to see millimetre details on façades and enable a photogrammetric reconstruction in the millimetre and centimetre range.
The panorama images were taken at point-to-point distances of 7-25m at the Campus Fallenbrunnen in Friedrichshafen.

3.2 Class Definition for Façade Component Instance Segmentation

The class definition began with researching façade datasets, instance segmentation, and standards for building description (see Chapter 2). Object classes proposed from CityGML and IFC's established standards were cross-compared to the related datasets, algorithms, and our raw image data taken. They resulted in only the TUM Façade dataset relying on the class definition of the standards. Despite those similarities, the TUM dataset is based on a LiDAR point cloud, and the raw data captured images from campus Fallenbrunnen allows for a higher detail level of façade reconstruction and component instance segmentation. Based on the circumstances, it was time to create a new classification for façade component instance segmentation based on images.
The TUM Façade proposal based on CityGML provided the template for the façade classes in images. In adaptation, the decision was made to inherit the object classes hierarchically to embed the object's context on the façade or building. For tracking, every object gets an ID. Fig. 1 is a class graph with the proposed inheritance order, light-grey coloured object classes for annotation, and additional object attributes for annotation. Table 2 contains the object classes and their relationship to their inheritance order, which begins with boundaries to label façades and roofs on buildings. This information should be used to extract the outer contours and surfaces in 3D reconstruction. Glass boundary annotations are used to annotate the glass

behaviour of reflection, mirroring, or transmission. All other parts for describing the façade as feature objects were proposed to distinguish 'additions' on the inheritance order. Construction-based additions are mainly based on cityGML 'BuildingInstallation' openable classes on the CityGML '_opening' feature. The opening subclasses were derived from standardized German 'Criteria for using windows and external doors by DIN EN 1435' terminology (DIN 18101:2014-08; DIN 18055:2020-09). The detail level of openable parts containing casement, handle, and inlays, like glass, could be annotated by boundaries. The mechanical moveable parts of buildings are considered in this much detail in order to reconstruct the building correctly by opening doors and identifying in which directions the parts are openable. Additionally, installations are mainly based on CityGML 'BuildingInstallationInt' and contain functional and technical installations around the house. The miscellaneous class contains objects like house decoration and non-technical communication used objects. All surrounding objects are distinguished as Elements.

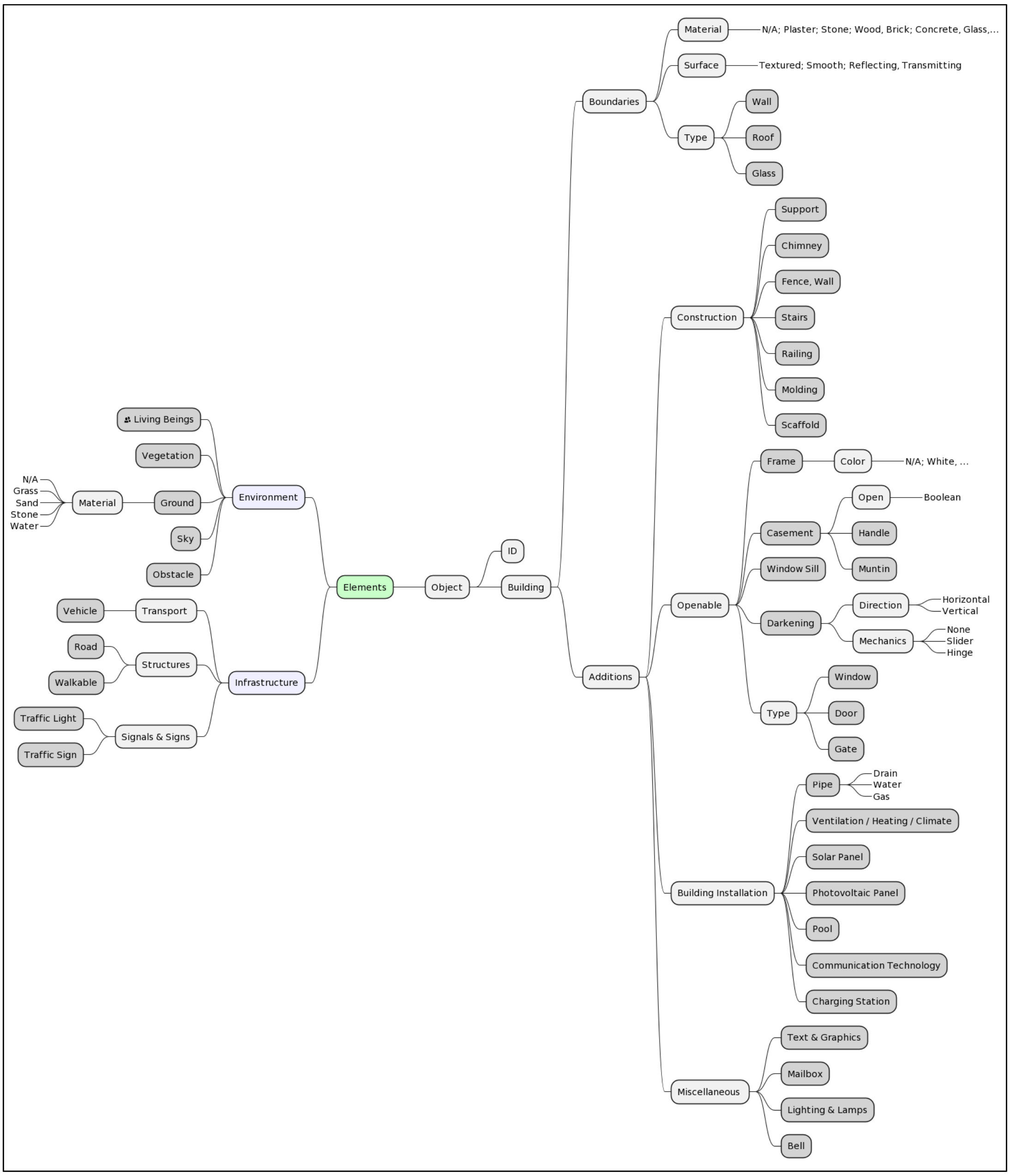


Fig. 1: Visualization of object classes and attributes in their relation and inheritance.

Tab. 2: Proposed object classes for annotation with inheritance order.

Inheritance Order	Proposed Classes for Annotation	Number of Classes
Object → Building → Boundaries → Type	Wall; Roof; Glass	3
Object → Building → Additions → Construction	Support; Chimney; Stairs; Fence; Stairs; Railing; Molding; Scaffold	8
Object → Building → Additions → Openable	Frame; Window sill; Darkening	3
Object → Building → Additions → Openable	Casement with Handle; Muntin	3
Object → Building → Additions → Openable → Type	Window; Door; Gate	3
Object → Building → Additions → Installations	Pipe; Ventilation/Cooling/Climate; Photovoltaik; Solar, Pool; Communication Technology; Charging station	7
Object → Building → Additions → Miscellaneous	Text & Graphics; Mailbox; Lighting & Lamps; Bell	4
Object → Elements → Environment	Living Beings; Vegetation; Sky; Obstacle	4
Object → Elements → Environment → Ground → Type	N/A; Grass; Sand; Stone; Water	5
Object → Elements → Infrastructure → Transport	Vehicle	1
Object → Elements → Infrastructure → Structures	Road; Walkable	2
Object → Elements → Infrastructure → Signals & Signs	Traffic Light; Traffic Light; Road Sign	3
Object → Elements → Infrastructure → Lightning	Streetlamp	1

4 Annotation

The annotation section of this paper separates the annotation guideline with rules from annotation examples based on the proposed classification.

4.1 Annotation Guideline

In the annotation process, ambiguities and ambiguity interpretations occur in images. To minimize these difficulties and to regulate the relations between the instantiated objects, the following guideline annotation is created:

1. Instances are annotated with polygons that have no holes.
2. Instances are allowed to lie on top of each other. Overlapping is to be avoided if possible.
3. If a surface or object is partially obscured by something in front of it and the alignment of the edges is recognizable, it is also marked in this area.
4. An openable instance, like a window or door, consists of at least one fixed frame and one or more movable casements.

The first and second rules show that the actual shape of some surfaces is only created by subtracting other instances. This can lead to additional work during reconstruction, which should be avoided. An example of the third rule is trees, which allow them to see partially through and identify the object behind them with its boundaries. The rule four addresses all physically openable, mechanical parts of a building that consist of a fixed frame and a moveable casement.

4.2 Annotation Praxis Examples

At this point, it should be mentioned that every panorama image is captured with a focal length of 16mm on a Sony IMX477 image sensor in 12 MP resolution. Every image is taken with the gimbal's longitude and latitude 5° steps, which resulted in 250 to 432 single pictures merged into a panorama image.

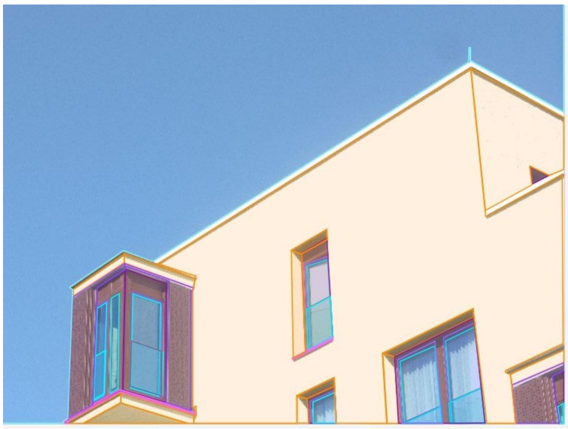
Manual image annotation for ground truth is done on single images with CVAT. The plan is to annotate 400 images and use transfer learning to annotate the other pictures taken of the Campus Fallenbrunnen to create a dataset. The dataset is a future annotation output and is not part of this paper. Therefore, data quality, quantity, transfer learning, and the usage of the data ground truth are no bearing in this paper.

This paper is about the classes and annotation, for which this section presents two examples of the annotation process.

Example 1:

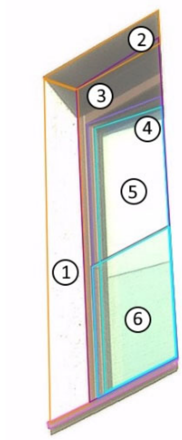
Fig. 2 shows an example of a façade boundary and openable object annotation. The left upper visualization (a) shows the extensive surface annotation of the sky (b.1) and wall boundary (b.2) instances. This large wall instance (b.2) is a reference according to the guidelines; it has no holes, allowing us to associate the window (b.3) with it. According to the guidelines, the windows on the bottom side (b.4) could be cut out because the wall does not enclose them in the image. The wall instance (b.2) is linked with an edge (b.5) to another wall instance (b.6).

The complexity of detailed window and glass surface annotation can be seen in (c). The depth-shifted wall surface boundaries (c.1-2) surround the window frame (c.3). The window has a hinge on the left side from the point of view and an opened casement (c.4) with a glass insert (c.5) which stretches from the top to bottom. This can be seen from the shade of the glass surface element (c.6). The top edge in front of the window is transmitting.

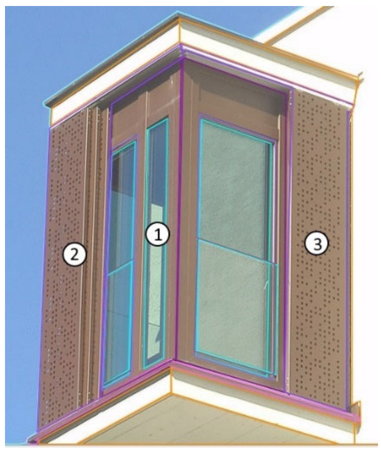


a. simplified annotation of the large surfaces (wall and sky)

b. Annotated wall surface polygon without hole and windows



c. The window consists of a frame and a casement with a glass pane.



d. The darkening partially covers the left window.

Fig. 2: First example annotation of boundaries, openable objects

Fig. 2.d has annotated object instances from (a) to (c), five wall surfaces, and the same schematic construction of windows as in (c). Nevertheless, window (d.1) has no handle seen by transmission and no protective glass downfall protection like the other two windows, so it is not openable but annotated as openable. The darkening (d.2) partially covers the window so that the left edge of the window frame, casement, and glass are not visible from this point of view. The darkening (d.3) has the exact dimensions of the left one, seen by counting dots, so there is an option to 3D reconstruct whole the window (d.2).

Example 2:
Example annotation two, the image annotation of the DHBW main building in Friedrichshafen, is shown in Fig. 3. According to the guideline boundaries, the image contains openable objects, building installations, miscellaneous, ground objects, and infrastructure objects. The annotation was done in the order of building (1), ground (2), stairs (3), building installation, and openable objects (4). Difficulties of annotation are described in the following.
The image is taken in a close parallel to the wall of the building (1) and shows annotation difficulties from this point of view: Annotation of the left two wall boundaries connected by an edge and the drainpipe is not a problem. However, interpreting the wall holes at the left wall from front to back without knowledge of other pictures or context embedding is complex. Below mark (1) is a darkening installation; further down is a windowsill on top of the wall. The windowsill is not in contact with the ground. Considering this instance under the context of the guideline, there must be an openable part, like a window. Conclusively, the windows going to the back have only a windowsill. The other effect while labelling the wall holes in the plane boundary is vanishing and making labelling challenging to the back.

Fig. 3: Second example annotation of a more complex outdoor scene

The inner courtyard near the building (2) is considered to have different ground surfaces (grass, stone, sand) and has some city furniture (benches, garbage cans, bicycle stands) marked as obstacles and vegetation (trees) on it. At the bottom of the image is a stair (3) marked with a wall in the background. The plaster ground and slabs are marked as walkable but not visible in the image to reduce coloured overlappings. The openable annotation of the window (4) is an example of partial through visibility and depth-stacked order of objects during annotation. The tree leaf density is so low that the regular contour of the window can be annotated, but the sill is not clearly visible and cannot be annotated. However, with the assumption that the same part is used for building construction, the windowsill should be the same as the other ones of the wall. Also time-consuming is the tree contour annotation caused by the many branches and leaves.

Both examples demonstrate the label process according to the proposed guideline and show the high detail level of façade component instance segmentation.

5 Conclusion

In our work, we presented an inherited and hierarchical object classification and annotation guideline that proposes a component instance segmentation in RGB façade images to reconstruct building façades in CityGML LOD3 (GRÖGER et al. 2021), IFC standard (ISO 16739-1 2018) and beyond.

Preliminary façade images in 12 MP resolution and exact global 6DoF pose and position of the campus Fallenbrunnen in Friedrichshafen were taken. These high-resolution and detailed images allow us to reach a higher detail level on façade instance segmentation.

The façade component instance segmentation classes proposed by CityGML and IFC's established standards were cross-compared to the related datasets, algorithms, and our raw image data. As a result, the TUM Façade LiDAR proposed classification was used as a template, and in adaption, the classes were hierarchically inherited according to CityGML and IFC. An extension of openable classes like doors and windows was done to a higher detail level, including frame, casement, and handle. Additionally, an extension of a subset of classes from CityGML 'IntBuildingInstallations', like Communication and Climate, was added for better usage, also context embedding object attributes were added. This results in an overview table and inheritance visualizing a class graph for annotation of the images.

For image annotation, a guideline was created to annotate the objects with closed polygons; overlapping or partially visible objects are allowed to be annotated, and openable objects consist of a frame and casement. According to this guideline, two images were example-wise annotated. The first example is a detailed scene with boundaries and openable objects. The second image was taken close to the wall at the inner of a U-shaped building, which makes embedding context necessary while simultaneously containing obscured objects, ground, vegetation, and building installations, all of which makes the annotation process more complex. Based on the practice annotation, the classes were iteratively improved.

Summarized, our proposal described the creation of classes, their hierarchical inheritance, and annotation for façade component instance segmentation based on images. In comparison the presented proposal is photogrammetric reconstruction-oriented like CFP (WANG et al. 2023) and standardized output oriented like TUM-Façade (WYSOCKI et al. 2022). The proposition closes the gap between its two predecessors by fulfilling both attributes as TUM-Façade did for LiDAR, but for façade image-based instance segmentation.

Based on the classes of this proposal, a dataset will be created to train in instance segmentation algorithms and photogrammetric façade reconstructions.

6 ACKNOWLEDGMENTS

A special thank you for support goes to Peter Buckel and Lucas Halt (DHBW Ravensburg, Technical Faculty) and Venus Shah (IWT Friedrichshafen) for data acquisition and annotation. This work was partially funded by the Bundesministerium für Bildung und Forschung (BMBF) through grant 01IS23037.

7 References

DIN 18055:2020-09, 2020: Criteria for the use of windows and external doors in accordance with DIN EN 14351-1. Berlin, Beuth Verlag GmbH, https://www.beuth.de/de/norm/din-18055/321515051, letzter Zugriff 28.01.2024.

DIN 18101:2014-08, 2014: Pedestrian Doors - Doors for buildings - Sizes of door leaves, position of hinges and lock - Interdependence of dimensions. Berlin, Beuth Verlag GmbH, https://www.beuth.de/de/norm/din-18101/204907014.

FRANK, F., HOEGNER, L., BUCKEL, PETER & DALM, KRIS, 2023: An Alternative Raw Data Acquisition Approach for Reconstruction of LOD3 Models. Recent Advances in 3D Geoinformation Science 2023, https://doi.org/10.1007/978-3-031-43699-4_29.

FRÖHLICH, B., RODNER, E., & DENZLER, J., 2010: A Fast Approach for Pixelwise Labeling of Facade Images, Istanbul, 20th International Conference on Pattern Recognition (ICPR 2010), http://dx.doi.org/10.1109/ICPR.2010.742.

GRIFFITHS, D. & BOEHM, J., 2019: SynthCity: A large scale synthetic point cloud. University College London, https://doi.org/10.48550/arXiv.1907.04758.

GRÖGER, G., KOLBE, T., NAGEL, C. & HÄFELE, K., 2021: OGC City Geography Markup Language (CityGML) Encoding Standard. Open Geospatial Consortium, https://mediatum.ub.tum.de/1145731, letzter Zugriff 28.01.2024.

ISO 16739-1, 2018: Industry Foundation Classes (IFC): Towards a Common Information Model for the Construction Industry: buildingSMART. International Organization for Standardization (ISO), https://www.iso.org/standard/70303.html, letzter Zugriff 28.01.2024.

KORČ, F. & FÖRSTNER, W., 2009: eTRIMS Image Database for Interpreting Images of Man-Made Scenes. University of Bonn, Department of Photogrammetry, https://www.ipb.uni-bonn.de/projects/etrims_db/, letzter Zugriff 28.01.2024.

LOTTE, R., HAALA, N.; KARPINA, M., ARAGÃO, L. & SHIMABUKURO, Y., 2018: 3D Façade Labeling over Complex Scenarios: A Case Study Using Convolutional Neural Network and Structure-From-Motion. Remote Sensing, **10**(9), 1435, https://doi.org/10.3390/rs10091435

LU, Y., WEI, W., LI, P., ZHONG, T., NONG, Y. & SHI, X, 2023: A deep learning method for building façade parsing utilizing improved SOLOv2 instance segmentation. Energy and Buildings, **295**, 113275, https://doi.org/10.1016/j.enbuild.2023.113275.

RIEMENSCHNEIDER, H., KRISPEL, U., THALLER, W., DONOSER, M., HAVEMANN, S., FELLNER, D. & BISCHOF, H., 2012: Irregular lattices for complex shape grammar facade parsing. IEEE Conference on Computer Vision and Pattern Recognition, https://doi.org/10.1109/CVPR.2012.6247857.

RIEMENSCHNEIDER, H., BÓDIS-SZOMORÚ, A.; WEISSENBERG, J. & VAN GOOL, L., 2014: Learning Where to Classify in Multi-view Semantic Segmentation. Computer Vision – ECCV 2014, https://doi.org/10.1007/978-3-319-10602-1_34.

SCHREINER, K., HANDEL, C., OBSTBAUM, M., BEUTENMÜLLER, F., MORICH, L., LANGER, M. & HARTH, M., 2021: Virtual Validation of Autonomous Driving Functions in Urban Areas with Mixed Traffic. 21. Internationales Stuttgarter Symposium. Proceedings. Springer Vieweg, Wiesbaden, https://doi.org/10.1007/978-3-658-33521-2_19.

SCHWAB B. & WYSOCKI O. 2021: LOD3 Road Space Models. https://github.com/savenow/LOD3-road-space-models, letzter Zugriff 28.01.2024.

TANG, Y., ZHOU, Y., YANG, K., ZHONG, Z., RAY, B., LIU, Y. ZHANG, P. & CHE, J., 2022: Automatic Map Generation for Autonomous Driving System Testing, https://doi.org/10.48550/arXiv.2206.09357.

TEBOUL, O., SIMON, L., KOUTSOURAKIS, P. & PARAGIOS, N., 2010: Segmentation of building facades using procedural shape priors. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, https://doi.org/10.1109/CVPR.2010.5540068.

TYLEČEK, R. & ŠÁRA, R., 2013: Spatial Pattern Templates for Recognition of Objects with Regular Structure. German Conference on Pattern Recognition, https://doi.org/10.1007/978-3-642-40602-7_39.

WANG, B., ZHANG, J., ZHANG, R., LI, Y., LI, L. & NAKASHIMA, Y., 2023: Improving Facade Parsing with Vision Transformers and Line Integration, https://arxiv.org/pdf/2309.15523v5.pdf, letzter Zugriff 28.01.2024.

WYSOCKI, O., HOEGNER, L. & STILLA, U., 2022: TUM-FAÇADE: Reviewing and enriching point cloud benchmarks for façade segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **XLVI-2**/W1-2022, 529-536, https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-529-2022.

WYSOCKI O., SCHWAB B. & WILLENBORG B., 2022: Awesome CityGML, https://github.com/OloOcki/awesome-citygml, letzter Zugriff 28.01.2024.