8TH INTERNATIONAL ISPRS WORKSHOP

LOWCOST 3D - SENSORS, ALGORITHMS, APPLICATIONS

LIÈGE université
GeoScITY

Automatic upgrade of 3D building models to LoD3 using 3D Point Clouds and Grounding DINO

Anass Yarroudh, Abderrazzaq Kharroubi, Imane Jeddoub, Roland Billen

GeoScITY, Geomatics Unit, UR SPHERES, University of Liège, Belgium

Friday, 13/12/2024: 10:00am - 10:15am

Università degli Studi di Brescia, Brescia, Italy

About the project

- 3D workflow automation for advanced simulation and training for rail and urban mobility solutions, or TrackGen, is a research and development project whose aim is to automate the creation of photo-realistic 3D railway simulation environments based on a set of heterogeneous geographical data.

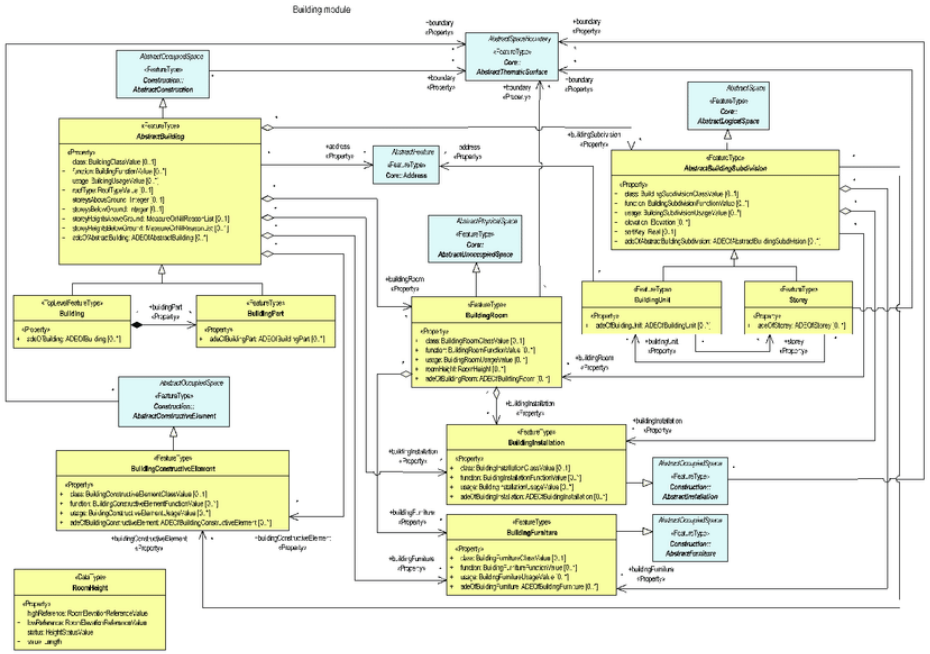
Plan

- Background
- Motivation and objectives
- Methodology
- Results and discussion

Buildings in 3D City Models

CityGML

CityJSON

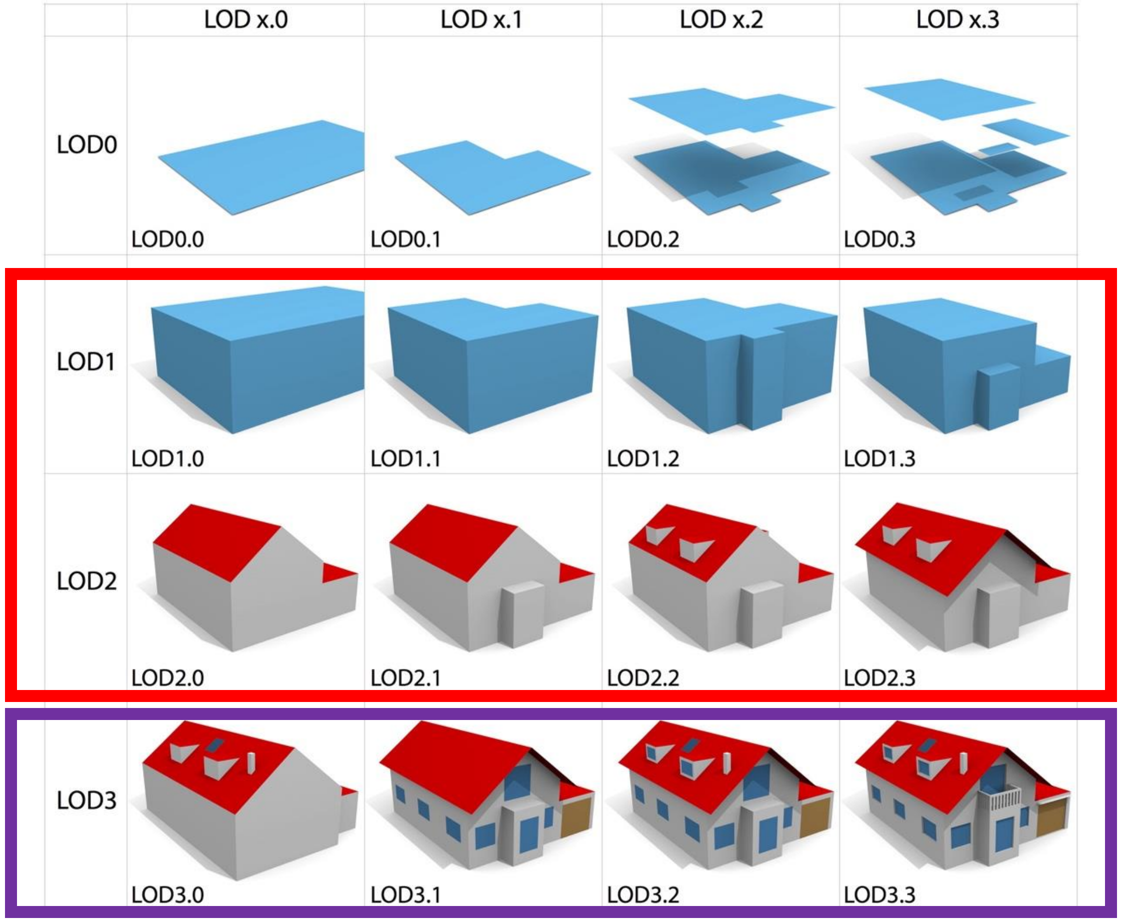


Level of Detail (LoD)

- Distinguish various representations of semantic 3D city models at different scales
- Level of the geometric detail within a model
- Biljecki et al. (2016) proposed an improved LoD specifications for 3D Building Models

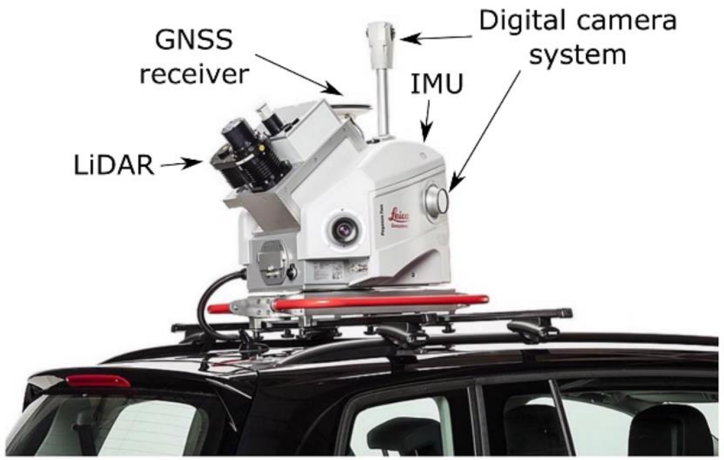
Which LoDs can be obtained using aerial LiDAR data and/or DSM?

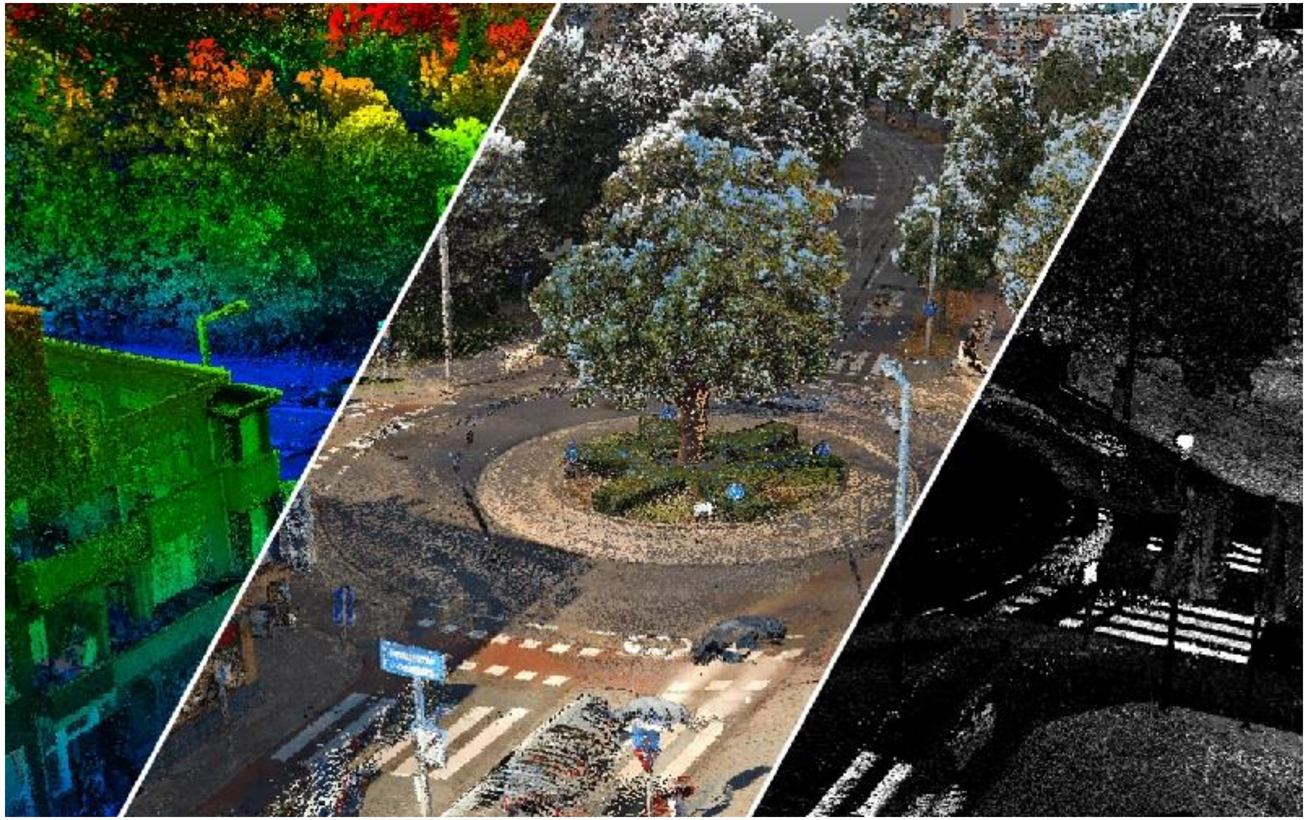
How can we upgrade these models to LoD3?



Biljecki, F., Ledoux, H., Stoter, J.: An improved LoD specification for 3D building models. Computers, Environment and Urban Systems, 59, 25–37 (2016)

Mobile Mapping Data





Grounding DINO

- Grounding DINO is a state-of-the-art model that combines visual grounding with detection, enabling it to localize objects in images based on textual queries, achieving strong performance even on open-vocabulary tasks.

- A zero-shot detector identifies and localizes objects in images without being explicitly trained on those object categories, using generalizable knowledge from pre-trained models or embeddings.

Problematic

How to use mobile mapping LiDAR data and Zero-shot object detection, i.e. Grounding DINO, can be used to upgrade LoD2 building models, resulting in detailed LoD3 models ?

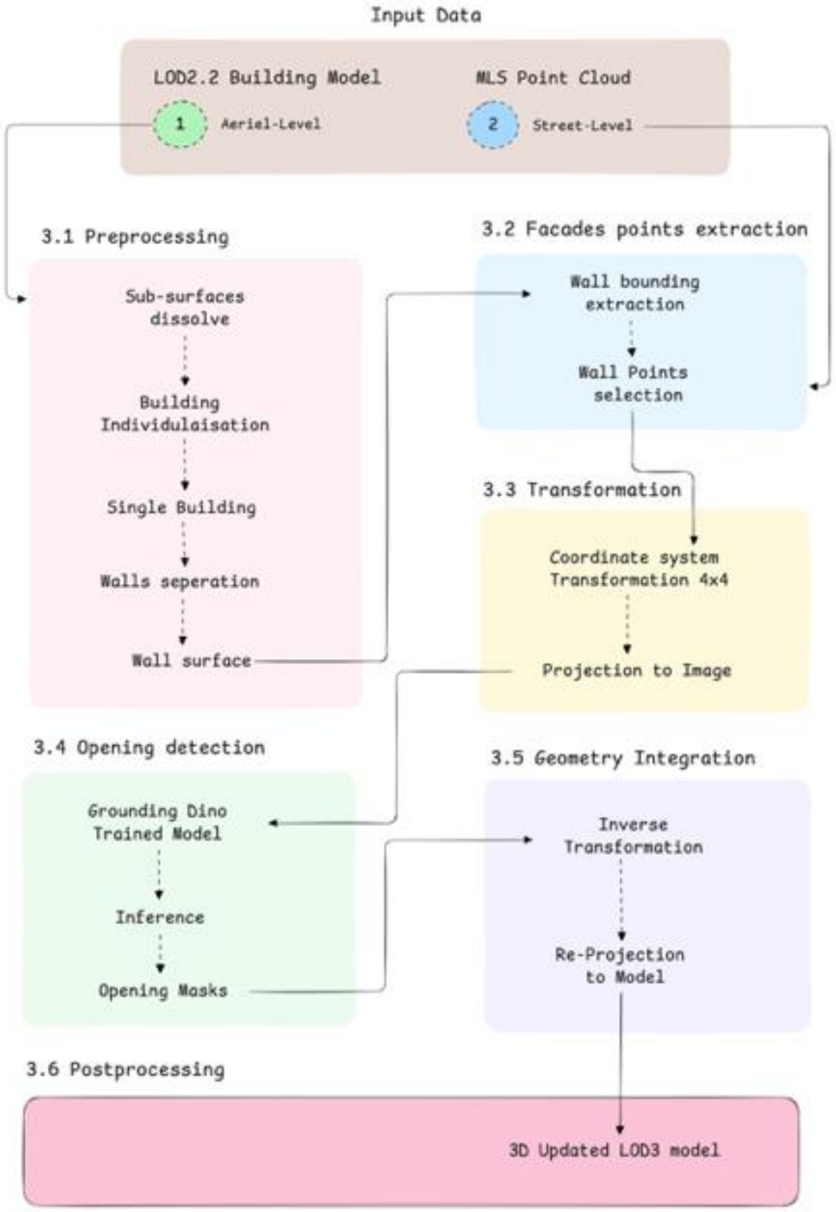
Objectives

- **Developing an approach for automatic upgrade of 3D building models to LoD3 using LiDAR data at the street level and zero-shot object detectors.**
- Data structuring and storage in accordance to existing 3D city models standards (CityGML 3.0 and its CityJSON encoding)

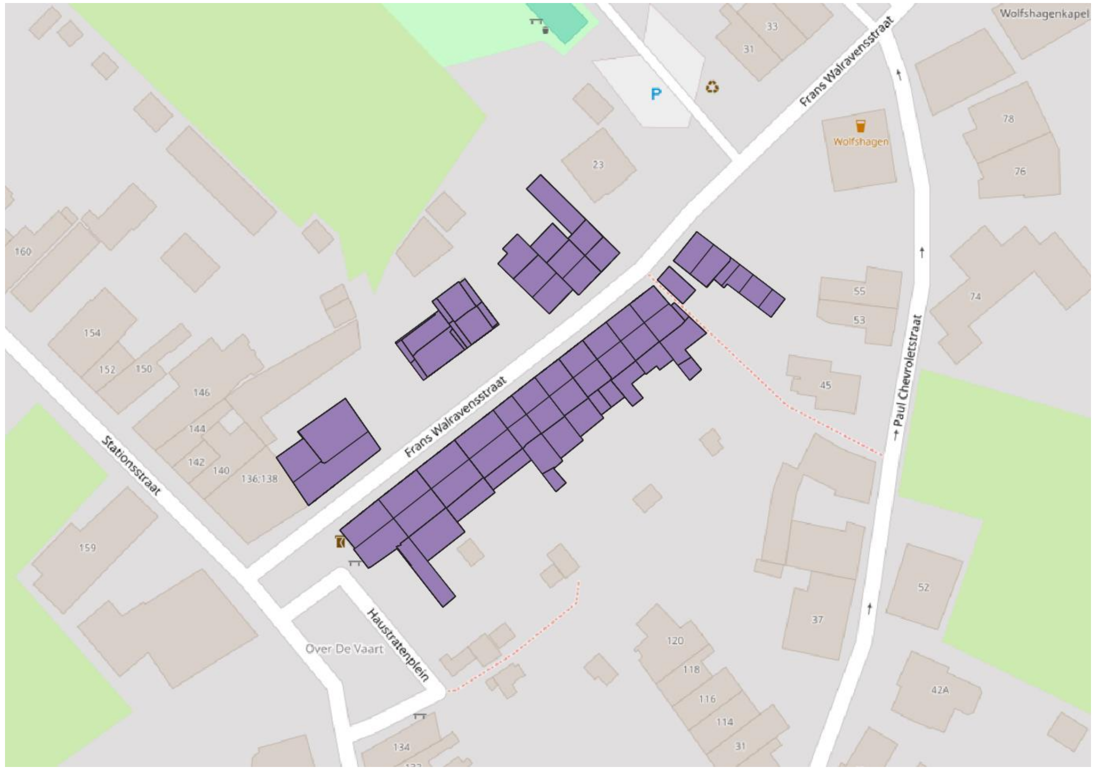
LIÈGE université
GeoScITY

Methodology

- ❑ Input: LoD2.2 Building Model + MLS Point Cloud
- ❑ Preprocessing: 3D models preparation for processing
- ❑ Façades points extraction from 3D MLS data
- ❑ Transformation to façade's plane (2D image)
- ❑ Opening detection using Grounding DINO
- ❑ Geometry integration/reprojection to model
- ❑ Output: 3D LoD3 building model



Testing data



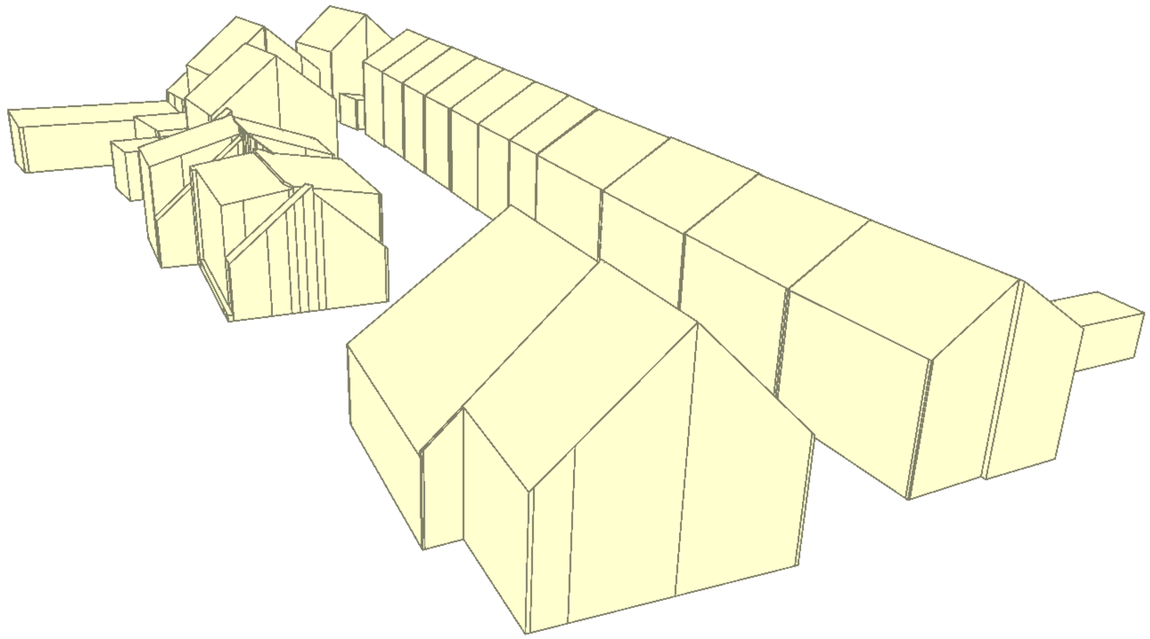
Sample of 18 building - LoD2.2 models generated from aerial LiDAR

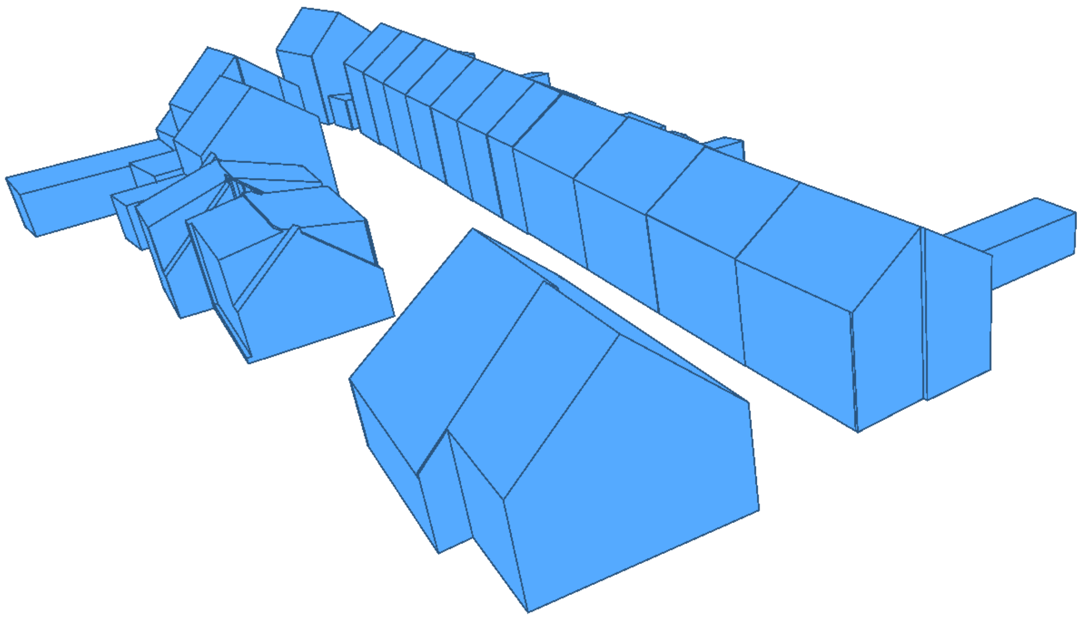
Corresponding 3D point cloud from Cyclomedia MLS data

Preprocessing

Wall Surface merging: In LoD2 model preprocessing, wall surfaces representing the same facade are merged based on topology (shared edges) and coplanarity (parallel normal vectors).

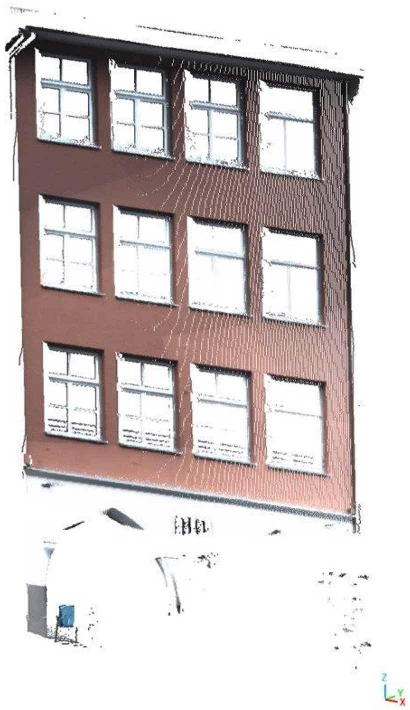
-> This merging ensures accurate segmentation of the 3D point cloud into facade sections, avoiding incomplete or missed detections of openings.



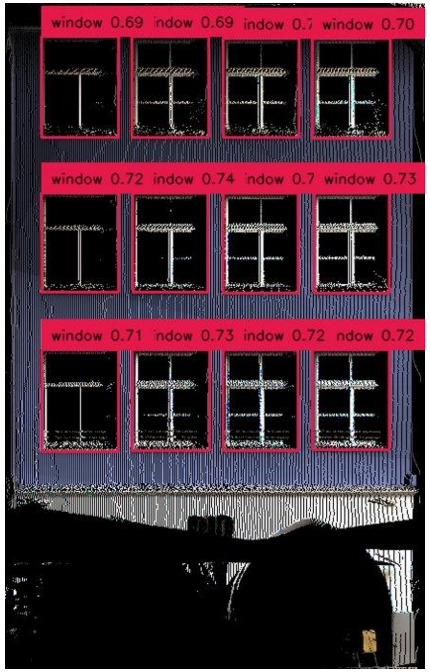


Detailed methodology

Prototype

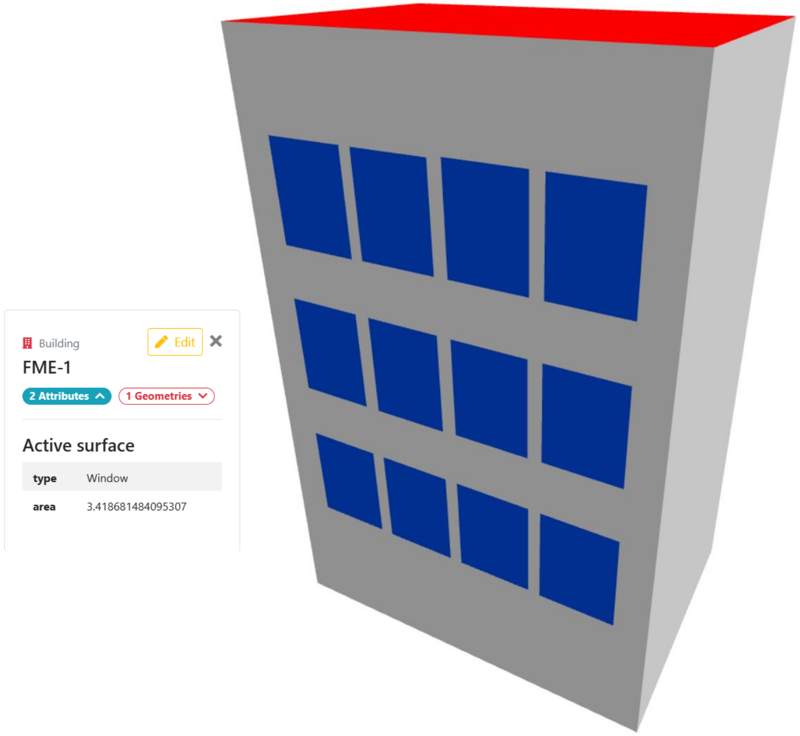


3D point cloud of façade obtained by instance segmentation of façades from TLS data



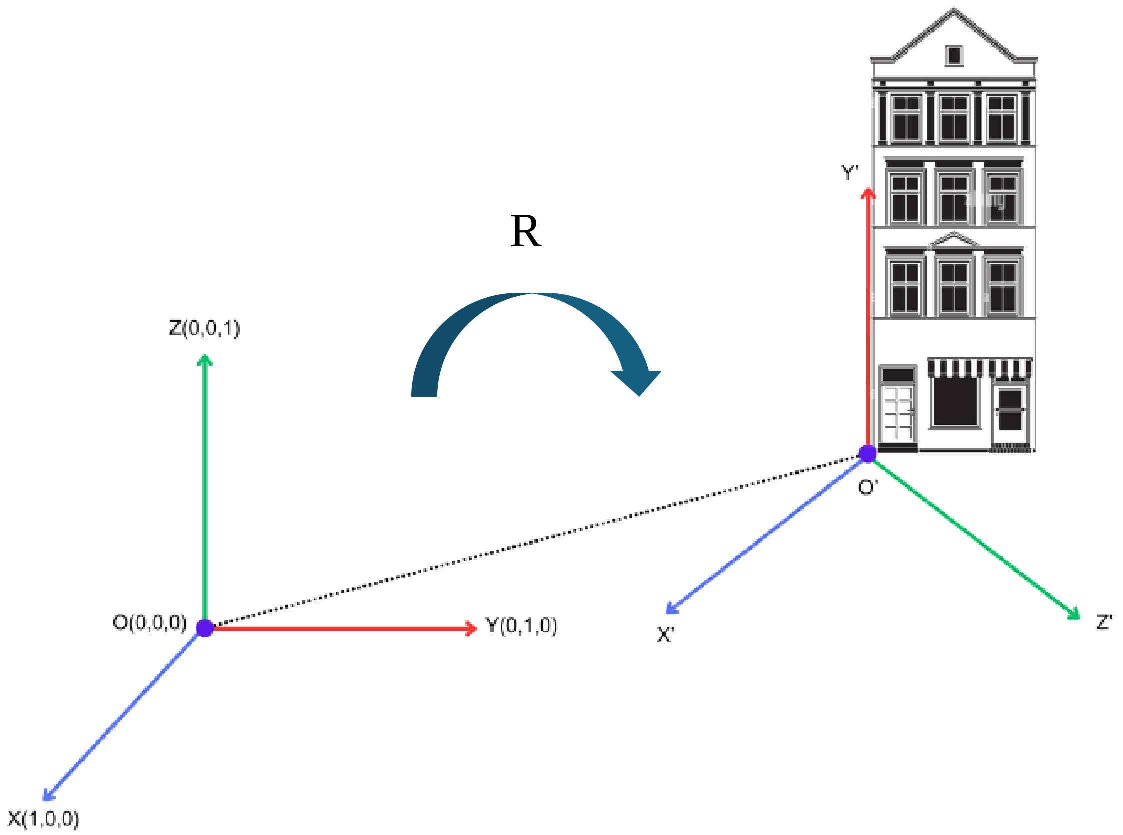
Projection into an image and object detection using GroundingDINO

Reprojection of masks values into point cloud using inverse transformation matrix



Geometry and semantics integration into LoD2.2 models
Results: LoD3 building model

Façades points extraction and transformation



Transformation of original points to a new local coordinates system defined on the facade

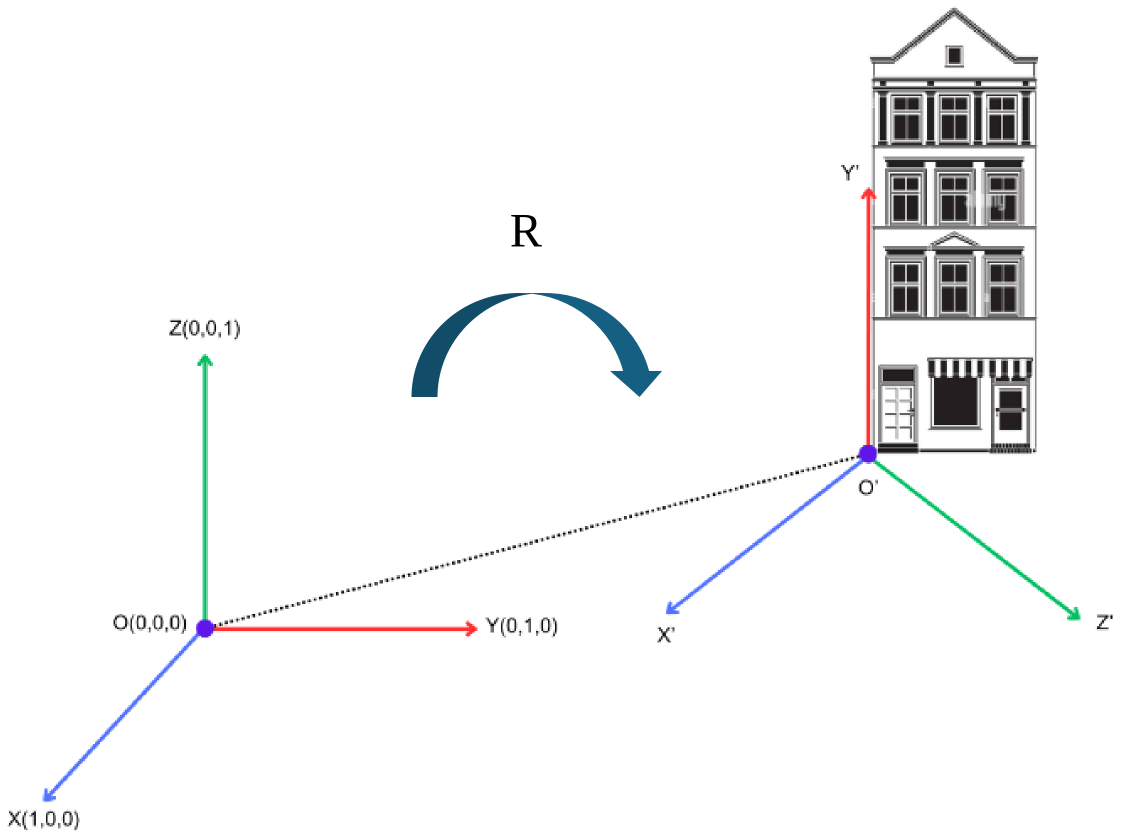
- The z-axis of the new coordinate system is aligned with the normal vector n.
- The x-axis is computed as the cross product of the normal vector n and the global z-axis:

$$X' = \frac{n \times Z}{|n \times Z|}$$

Where X' is the x-axis of the new coordinate system and Z is the global coordinate system's z-axis.

- The y-axis is determined by the cross product of the x-axis and the z-axis.

Façades points extraction and transformation



Transformation of original points to a new local coordinates system defined on the facade

The full transformation matrix M is then given by:

$$M = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

Where R is the rotation matrix and t is the translation vector:

$$R = R_x(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma)$$

$$t = O' - O$$

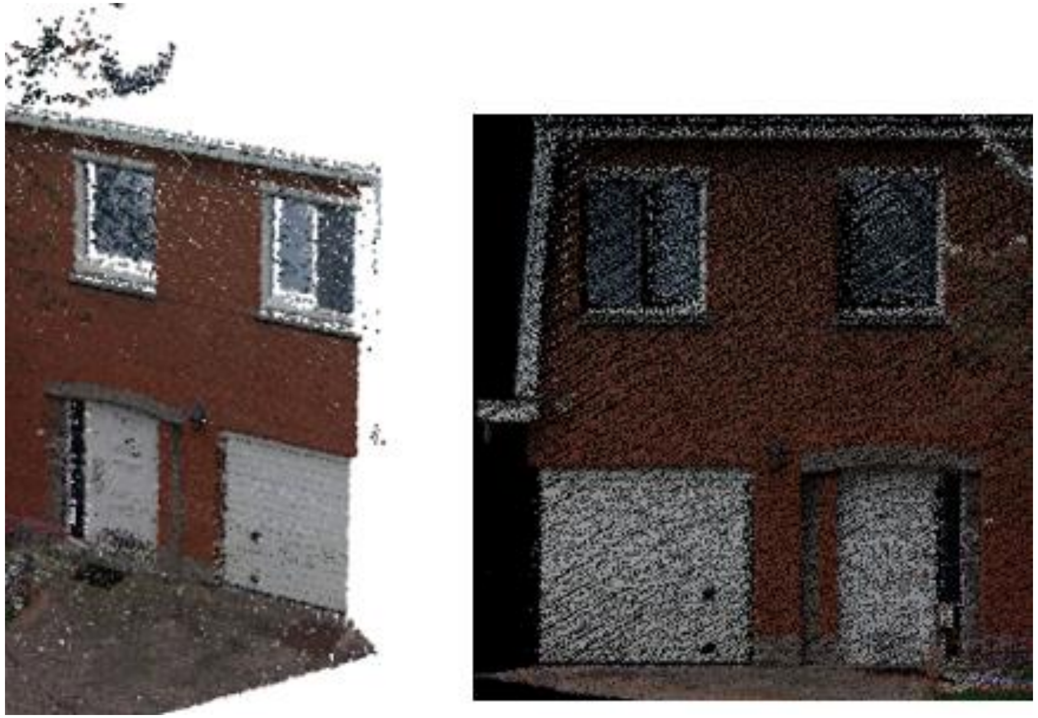
α, β and γ are rotation angles around the x-axis, y-axis and z-axis respectively.

The transformation of a point from the global to the local coordinate system is performed as follows:

$$\boldsymbol{p'} = \mathbf{M} \, . \, \mathbf{p}$$

p = (x, y, z,1) are the homogeneous coordinates of the point in the global coordinate system, and p′ = (x′,y′,z′,1) are the homogenous local coordinates.

Façades points extraction and transformation



Example of generated image (right) from 3D point cloud of facade (left).

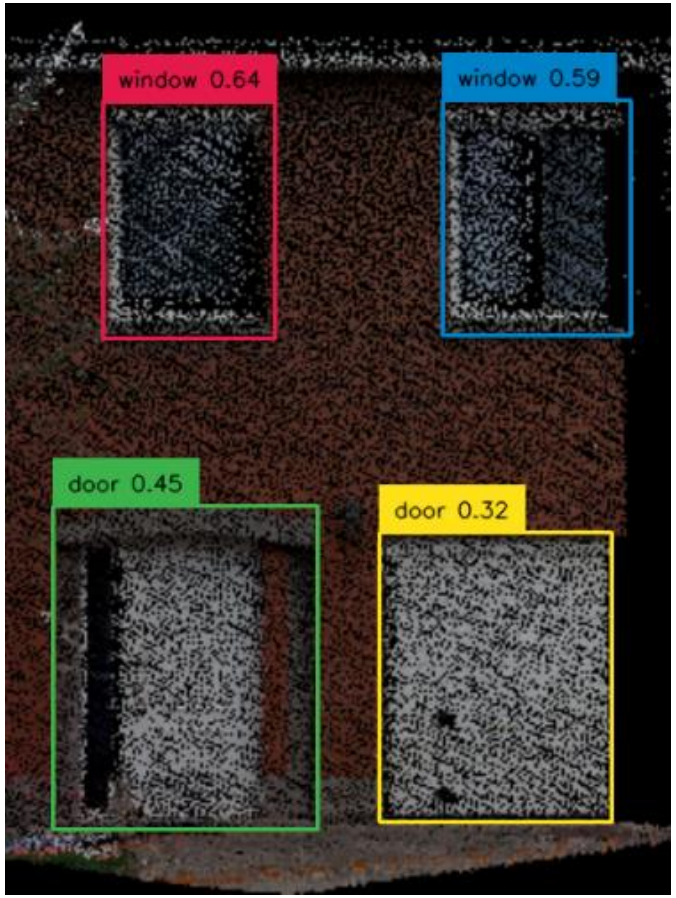
Once the points are transformed into the local coordinate system, they are then projected onto a 2D plane by mapping their x and y coordinates to pixel coordinates. The resolution σ of this mapping is defined to control granularity of the image. The pixel coordinates (u, v) are calculated as follows:

$$u = \frac{x' - \min(x')}{\sigma}$$

$$v = \frac{\max(y') - y'}{\sigma}$$

The result is a 2D image array where each pixel corresponds to a point, or a set of points, in the wall surface's plane, with its color determined by the original point cloud data. In case two points have the same pixel coordinates, the pixel is assigned the color value of the farthest point, i.e. the one with the highest z' coordinate.

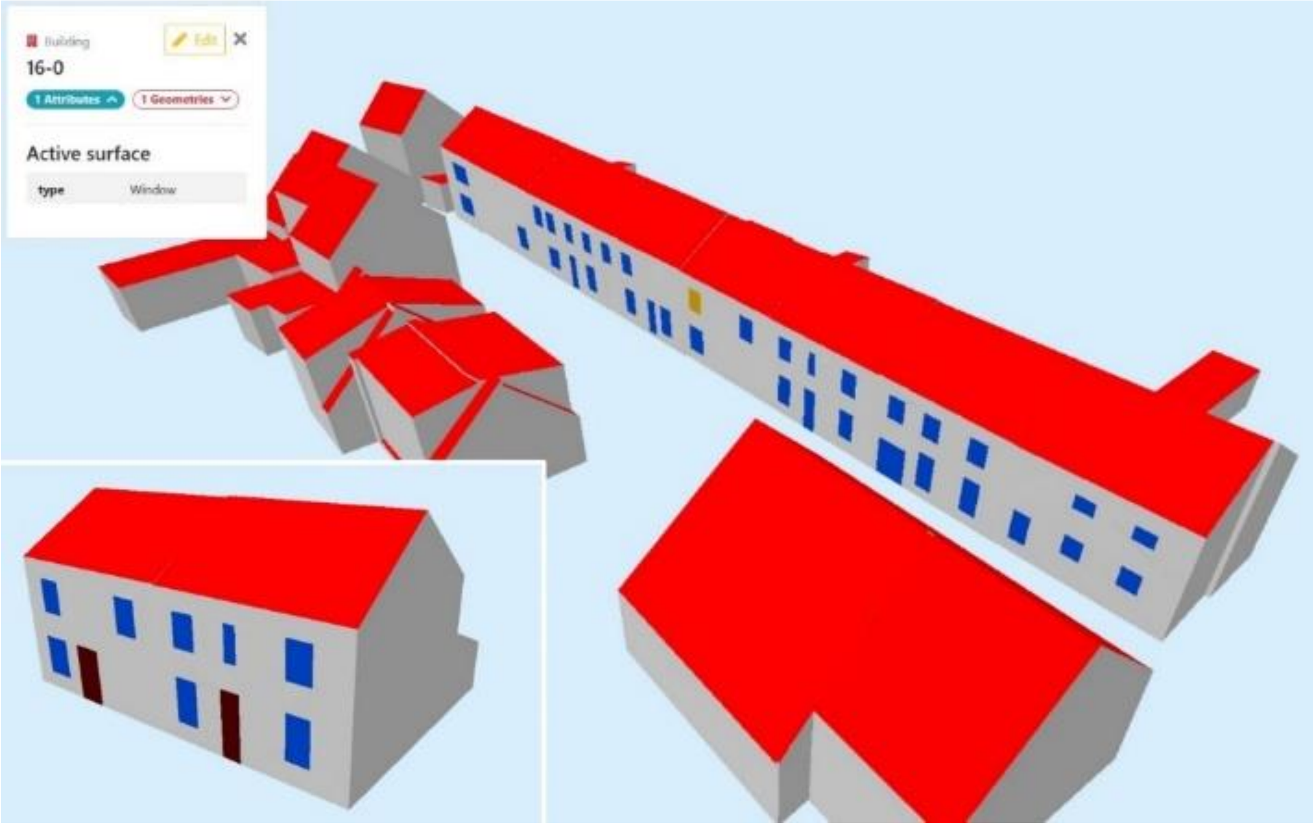
Openings detection and integration into models



- **Opening Detection**: Façade points are projected into 2D images for detecting windows and doors using the **Grounding DINO** model, which combines language and vision. Detected features are then **re-projected onto the global coordinate system**, using the inverse transformation matrix.

- **Geometry Integration and refinement**: Openings are created in wall surface polygons, and the geometries of detected elements are integrated and aligned with the façade's plane normal vector. A **postprocessing step** refines the model using **elevation** (to distinguish windows from doors) and **area** (to filter false detections).

Results and discussion

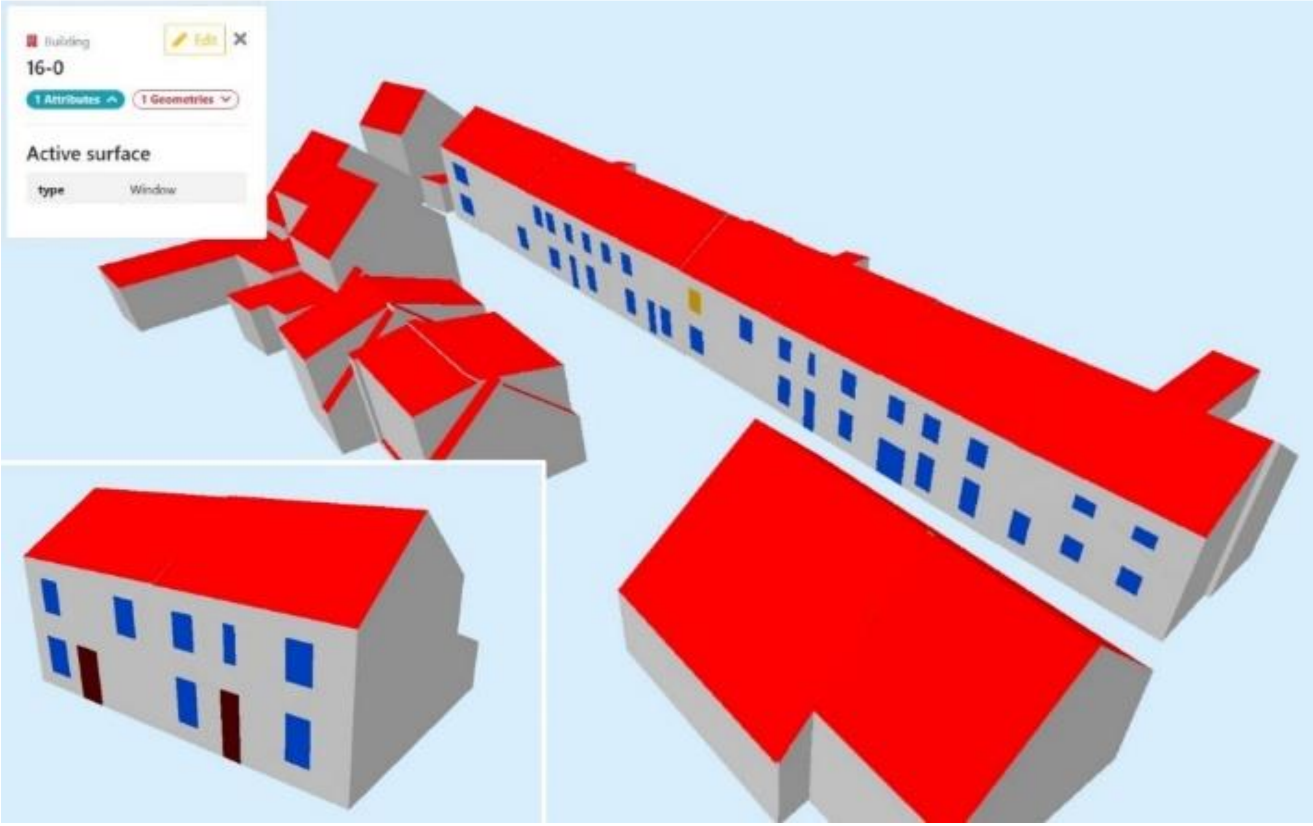


Detection performance

- Grounding DINO achieved a **detection rate of 61.53%** on a test dataset of **65 windows** from MLS colorized point cloud data (Belgium).
- Detected features were correctly mapped to the 3D models, ensuring proper **orientation** and alignment with the façade wall normal vector.
- Results were written in **CityJSON format** and inspected using Ninja Viewer.

Limitations

- **False Positives**: Caused by occlusions from nearby objects (e.g., cars, trees) in dense areas.
- **Point Cloud Density**: Lower density reduced detection accuracy and completeness, particularly for **smaller features** and higher floors.
- **Capture Conditions**: Detection quality was affected by the **angle and distance** of data capture, with oblique angles and larger distances leading to fewer façade points.



Recommendations

- Adjust **box and text thresholds** based on the 3D point cloud's completeness.
- Improve **data acquisition planning** to minimize occlusions and ensure sufficient point cloud density, especially for higher façades.
- Annotate 3D point cloud data to train 3D objects detectors.

LIÈGE université
GeoScITY

Conclusions and perspectives

Conclusions and perspectives

- We developed a methododology to upgrade LoD2.2 models to LoD3 using 3D point clouds and the Grounding DINO model for automated façade detection.
- The approach uses colorized point clouds instead of mobile mapping images, removing the need for image selection, and is scalable for large urban areas.
- Results depend on point cloud quality; low density and occlusions (e.g., cars, trees) can cause false positives or missed elements.
- Detection quality is affected by capture angles and distance, especially for higher floors.
- The lack of annotated datasets and the 3D-to-image transformation step can add challenges and computational costs to the method.
- Future work should focus on improving data acquisition, expanding training datasets, and optimizing processing for large-scale use.

8TH INTERNATIONAL ISPRS WORKSHOP

LOWCOST 3D - SENSORS, ALGORITHMS, APPLICATIONS

Automatic upgrade of 3D building models to LoD3 using 3D Point Clouds and Grounding DINO

Anass Yarroudh, Abderrazzaq Kharroubi, Imane Jeddoub, Roland Billen

GeoScITY, Geomatics Unit, UR SPHERES, University of Liège, Belgium

Friday, 13/12/2024: 10:00am - 10:15am

Università degli Studi di Brescia, Brescia, Italy