AGILE: GIScience Series, 7, 8, 2026. https://doi.org/10.5194/agile-giss-7-8-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

Maximizing Data Coverage: Fusing Street View and Oblique Imagery to Quantify Vertical Greenery Potential in Urban Areas

Aruscha Kramm [1], André Ludwig[2], and Bogdan Franczyk[1,2]

[1]Leipzig University, Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig
[2]Leipzig University, Information Systems Institute

Correspondence: Aruscha Kramm (aruscha.kramm@uni-leipzig.de)

Abstract. As urbanization intensifies globally, the Urban Heat Island (UHI) effect has emerged as a critical environmental challenge, inducing higher energy demands, compromised air quality, and significant public health risks. Vertical Greenery Systems (VGS) such as green facades and living walls, offer a spatially efficient adaptation strategy by utilizing vertical surface areas of the built environment for thermal regulation and microclimatic improvement, yet the absence of data-integrative methods hinders large-scale evaluation of factors defining a surface's suitability for VGS. Previous methodologies for estimating city-wide greening potential have successfully integrated semantic 3D city models (LoD2) with Street View Imagery (SVI) to derive key suitability factors such as Window-to-Wall Ratio (WWR) and Solid Wall Area (SWA). However, these approaches are inherently limited by the sparse spatial coverage of SVI, which is restricted to navigable road networks, leaving rear facades and inner courtyards unassessed, and which is frequently obstructed by foreground occlusion, leading to errors in the factor calculation. This consecutive work introduces a robust computational method that enhances the existing LoD2-SVI pipeline with Oblique Aerial Imagery extending the potential estimation to over 90% of the urban building stock. To mitigate occlusion and resolution disparities, we propose a multi-view fusion algorithm that aggregates detections across multiple views within one perspective and further across two perspectives. Our evaluation demonstrates that both data sources deliver comparable results when assessing identical facades. Further, our fusion approach significantly reduces systematic biases found in single-source estimations. Ultimately, while the fusion approach maximizes assessment reliability for walls with dual coverage, the integration of oblique imagery remains critical for scalability. Although it yields lower feature fidelity than street view, it provides the only viable means to assess surfaces lying beyond the navigable road network.

Submission Type. Algorithm, Case Study, Analysis

BoK Concepts. [GC] Geocomputation, [DM5] Modelling 3D, temporal and uncertain phenomena, [IP3] Image understanding

Keywords. oblique images, street view, vertical greenery, smart city, climate change, applied artificial intelligence

1 Introduction

The continuous densification of metropolitan areas has led to a significant replacement of natural, porous surfaces with impervious materials such as concrete and asphalt. This transformation contributes to the Urban Heat Island (UHI) effect, where cities exhibit markedly higher temperatures than their rural surrounding areas. The consequences are multifold: increased mortality during heatwaves, elevated energy consumption for air conditioning, which further expels waste heat into the urban canopy and a degradation of overall urban livability (Oliveira et al., 2022; Heaviside et al., 2017; Santamouris, 2014).

In this context, the integration of vegetation into the built environment is no longer merely an aesthetic aspiration but a functional necessity for climate adaptation. Urban greening can be achieved through horizontal interventions like green roofs and unsealing pavements, as well as by implementing greening on existing vertical surfaces such as building facades. While roofs present between 20 to 25% of the exposed urban area in dense cities (Santamouris, 2014), wall surfaces exceed the roof area by a factor of two to three (Köhler, 2008; Kramm et al., 2026). To leverage this extensive surface availability, this study focuses on Vertical Greenery Systems (VGS), which include green facades, living walls, and bio-curtains to reintroduce vegetation into dense urban areas. They offer a potent strategy to mitigate environmental stressors without competing for scarce ground-level space. Benefits from VGS include increasing the proportion of green spaces, reducing the temperature of building interiors in hot periods as well as the

surrounding micro-climate (Safikhani et al., 2014; Perez et al., 2014).

However, determining the suitability of a specific wall for VGS is a complex challenge. Conventional assessments rely on expert diagnostics, a process that lacks the computational scalability required for city-wide analysis. Relevant factors include geometric orientation and Solid Wall Area (SWA), alongside structural variables such as material composition and surface degradation (BuGG Bundesverband GebäudeGrün e. V., 2024).

While the theoretical benefits of VGS are well-established, there remains a notable gap in literature regarding computational methods for assessing the suitability of a wall regarding the mentioned factors. Given the data complexity, this study prioritizes parameters with the highest impact on UHI mitigation potential. Specifically, the Window-to-Wall Ratio (WWR) is a critical determinant as walls with multiple windows are generally less suitable due to higher maintenance costs and conflict with daylighting needs. Consequently, the SWA, representing the available uninterrupted masonry, is derived as a critical metric for greening capacity.

This paper is an advancement of the work of (Kramm et al., 2026), in which a foundational methodology for automating this assessment was established. By coupling semantic LoD2 city models with street view imagery (SVI), the study demonstrated that visual data is essential for deriving wall-specific WWR and SWA, as LoD2 models typically lack facade semantic information, modeling walls as blank polygons. However, reliance on SVI introduces systematic biases and coverage gaps. SVI is constrained to the navigable road network while further only being able to capture street facing facades (Gaw et al., 2022). This leaves a majority of vertical surfaces, such as rear facades and side walls, completely unassessed. These omitted surfaces are often those with the higher potential for greening: they are frequently less fenestrated than street facades and are less subject to architectural preservation constraints. In addition, street view images suffer from occlusions such as vegetation or vehicles.

To address the limitations of ground-based imaging, this study integrates Oblique Aerial Imagery. Oblique imagery is acquired from cameras facing four cardinal directions (North, South, East, West) and tilted at an angle (typically $40°$ to $50°$) relative to the nadir. Because oblique sensors are able to capture vertical surfaces including those inaccessible to terrestrial vehicles such as enclosed courtyards, the overall visibility of building walls can be increased, eliminating the spatial bias of road-network dependency (Gaw et al., 2022). However, the integration of oblique imagery introduces difficulties not present in SVI processing. These include significant geometric distortions due to perspective projection and lower resolutions. Further, oblique imagery is not immune to occlusions as taller buildings block the view of adjacent lower facades (Gaw et al., 2022).

To overcome the limitations of one-to-one mapping where a wall's potential is derived from only a single image, this work develops a multi-source framework fusing LoD2 geometry, SVI, and Oblique Aerial Imagery. The facade parsing pipeline is evolved to implement a multi-view fusion strategy: for each target surface, up to three images per source are analyzed. By aggregating detections from multiple viewpoints, the framework remains robust even when a surface is only partially visible in one image or only visible in one source. In the absence of an annotated ground-truth, this study evaluates the reliability of factor values acquired from oblique images. By cross-validating these results against SVI-derived values and performing a qualitative manual inspection, we quantify the extent to which oblique imagery can serve as a dependable extension for city-wide VGS potential assessment. A potential index is created and updated incorporating the newly assessed hidden facades, thereby refining the total estimated vertical greenery potential of the city.

The rest of the study is structured as follows: Section 2 highlights related work in the areas of automated calculation of VGS factors and multi-view fusion. In Section 3 the factors and index calculation are presented, while Section 4 explains the data used and methods implemented. The results are evaluated in Section 5 before the study is concluded in Section 6.

2 Related work

There is a notable lack of computational methods for estimating the overall potential of individual walls for vertical greenery. While existing research addresses individual factors such as Window-to-Wall Ratio (WWR), sunlight, or material, only one current method combines these relevant factors to assess a specific wall's total potential.

Factor computation. Although LoD2 data is standard for building-level analysis (Xu et al., 2023), it has rarely been utilized for single-wall surface analysis. However, it has been used to calculate singular factors needed for the vertical greenery potential. Biljecki et al. examined the error rate when estimating the ground surface area from LoD data, demonstrating that LoD2 data reduces surface area estimation errors to a range of <1% to 12%, compared to the higher error rates of LoD1 (Biljecki et al., 2016). Furthermore, Koehler estimated that the solid wall area in cities is roughly double the building ground surface, suggesting high potential for vertical greenery, though the methodology for this estimate remains simplified (Köhler, 2008). In a first attempt to computationally derive factors for greening potential, (Kramm et al., 2026) use LoD2 data in combination with SVI to calculate factors such as WWR and SWA.

Facade parsing. To accurately derive factors such as WWR and material from street view images, facade parsing is essential. Recent advancements in this field include symmetry-aware approaches, such as those by Liu et al.

(Liu et al., 2020) and Liu et al. (Liu et al., 2022), which incorporate symmetry into loss functions and model architectures, respectively. Other researchers focus on spatial dependencies; Sun et al. combine region proposals with Convolutional Neural Networks (CNN) (Sun et al., 2022), while Ma et al. utilize large kernels for long-distance relationships (Ma et al., 2021), although Lu et al. note the latter struggles with irregular facades (Lu et al., 2023). Regarding specific facade element detection, Kramm et al. successfully applied detectron2 for residential buildings (Kramm et al., 2023), while Sezen et al. established benchmarks using YOLO and R-CNN models (Sezen et al., 2022). Accurate facade parsing requires prior image rectification, which remains a challenge. Most methods rely on computer vision without leveraging camera data.

Multi-view fusion and geolocalization. To mitigate the limitations of single-perspective analysis, recent research leverages cross-view geolocalization and multi-view fusion to synthesize data from diverse viewpoints. Wegner et al. demonstrate the effectiveness of combining aerial and street-level imagery, using the former for coarse object localization and the latter for fine-grained classification (Wegner et al., 2016). To determine the position of objects that appear in multiple images, Nassar et al. employ Graph Neural Networks (GNN) to cluster detections into unique geographic coordinates based on spatial proximity. The GNN learns to cut edges between false matches (Nassar et al., 2020). Furthermore, multi-sensor pipelines have been proposed to enhance positioning accuracy. For instance, Krylov et al. fuse street-level images with airborne LiDAR. Segmentation in street-level images generates candidate objects and their position is estimated using monocular depth estimation. Candidate positions are then computed from LiDAR using Markov Random Fields (MRF) (Krylov and Dahyot, 2018). These approaches highlight a transition from isolated image processing to integrated, multi-source frameworks that provide a more robust and spatially consistent urban inventory.

3 Factor Definition and Index Construction

The selection, definition and calculation of factors for estimating vertical greenery potential in this work remain consistent as described in (Kramm et al., 2026). While factors such as wind exposure remain outside the scope of current data availability, this study focuses on factors that either determine a wall's initial suitability and are therefore indirectly required for the index, and evaluative factors, which directly populate the potential index.

Factors consisting of geometric attributes including *orientation* (angle-to-north), *cardinal direction*, *height*, *outside exposure*, and *ground-boundness*, are extracted directly from the LoD2 data. Of these, height, outside exposure and ground-boundness determine initial suitability. In order for a wall to be attributed as ground-bound, it needs to have ground-level contact. Walls that are not ground-bound are not taken into account. Outside exposure requires a wall to have no direct adjacent neighbours so that it is visible to the recording cameras.

To quantify available greening space, we derive the WWR by applying a CNN-based object detection pipeline to rectified images. This allows for the calculation of the SWA, representing the uninterrupted masonry surface and therefore the area available for greening.

For eligible walls, the potential index is computed following the logic established in (Kramm et al., 2026), prioritizing heat stress reduction. The index is weighted based on three critical components (Bustami et al., 2018):

- **SWA**: High values indicate greater greening capacity.
- **WWR**: Lower ratios denote fewer conflicts (e.g., daylighting needs or maintenance costs).
- **Solar Orientation**: South-facing surfaces are prioritized due to higher solar radiation, maximizing the cooling benefits of evapotranspiration and supporting optimal plant growth.

Walls are subsequently ranked according to this index, with a rank of 1 identifying the surface where greening offers the highest potential for heat stress mitigation.

4 Methods

Using a combination of LoD2 data and two different image sources, this approach aims to find facade elements in images of individual buildings' walls to derive the factors mentioned in section 3. To do so, the existing pipeline (Kramm et al., 2026) for LoD2 data processing and street view image extraction has been extended (see Figure 1). The following section illuminates data acquisition and preprocessing before explaining factor and potential index calculation.

4.1 Data sources

The LoD2 model is openly available (see Section 7) and was preprocessed following the mentioned preliminary approach. To summarize it briefly, the LoD2 data is used to source the geometry and metadata for approximately 1,300 buildings in the suburb Wiesdorf serving as test area in the city of Leverkusen (see Table 1). At this point, a model with higher level of detail is unavailable, such as LoD3, which would include facade openings. A significant challenge with this dataset is that buildings are often modeled as composite structures, consisting of a main building alongside smaller extensions. Consequently, outward-facing walls are frequently represented as multiple fragmented surfaces rather than single continuous units (see Figure 2 a) and b)), which requires extensive preprocessing to accurately estimate the potential for vertical greenery.

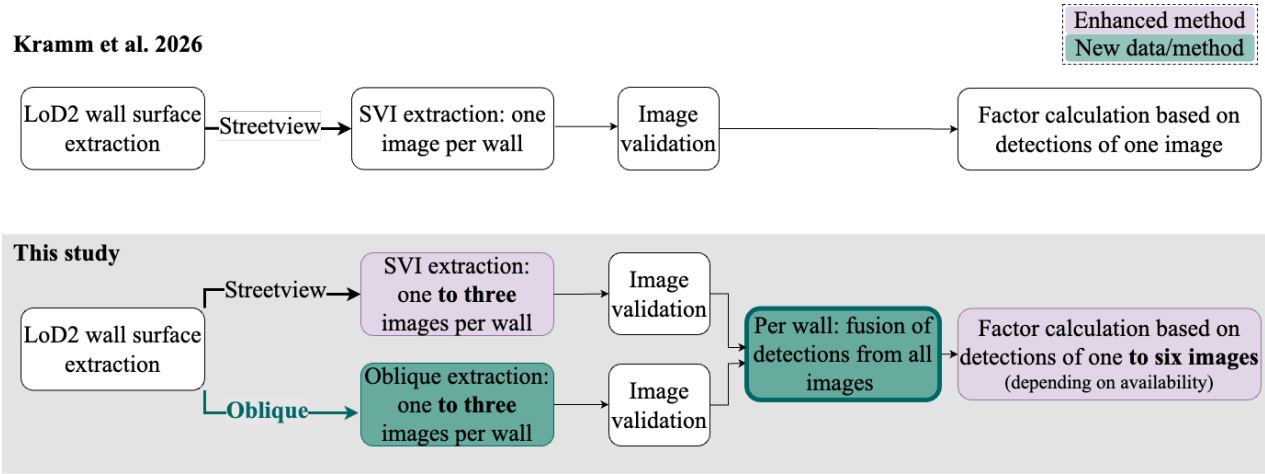
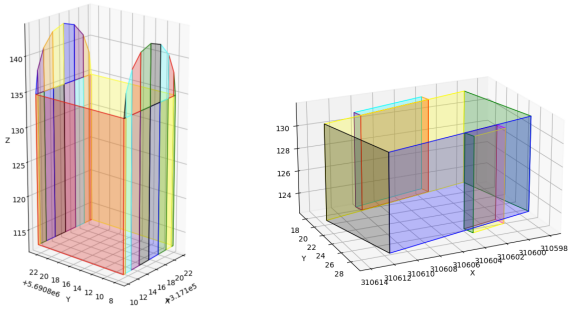


Figure 1. Methodological Evolution. Comparison of the baseline approach (Kramm et al., 2026) using single-source SVI (top) versus our proposed multi-view fusion framework (bottom). Our workflow integrates oblique imagery (green) and enhances the extraction logic (purple) to fuse up to six views per surface, significantly increasing robustness.



(a) (b)

Figure 2. LoD2 data: Fragmentedly modelled building (a) and building with extensions modelled as separate parts (b).

The preprocessing pipeline first focuses on data cleanup by removing invalid geometries. This includes eliminating surfaces defined by three or fewer coordinates, deleting surfaces with an area of 0 m^2, and filtering out non-vertical walls where the normal vector deviates by more than 5 cm. Following this, the process addresses the fragmented modeling within the LoD2 data, where single physical walls often appear as separate or insignificantly small surfaces. To resolve this, surfaces belonging to the same wall are merged into a cohesive unit, and any remaining surfaces that do not connect to the ground are removed from the dataset. For further processing, the dataset contains wall surfaces and ground surfaces of each building modeled as tuples of real world coordinates (x, y, z). **360° RGB panoramic images** of building facades in Leverkusen were retrieved for ground-level analysis via the *StreetSmart* API[1]. To do so, we were provided with credentials by the city of Leverkusen. Building facades had to meet two selection criteria, which serve to minimize the error rate for object detection: a maximum camera-to-wall distance of 20m and a minimum facade area of 5m^2. Without the camera-to-wall distance threshold, far-away building walls are fetched, on which facade details cannot be recognized. The minimum facade area ensures that fragmented walls mentioned above, for which wall merging did not create a cohesive unit, are not extracted, as object detection fails on these.

For qualifying facades, we identified candidate recording points characterized by optimal viewing angles and minimal occlusion by other buildings. In an evolution of the established street view extraction pipeline (Kramm et al., 2026), we retrieved three separate images per facade from varying positions, replacing the single-view approach to potentially mitigate ground-level occlusions. To derive factors such as the Window-to-Wall Ratio and the Solid Wall Area, facade elements will be parsed using object detection. To obtain an accurate result using these techniques, the facade image is rectified before processing to allow the method to recognize facade elements as precisely as possible. To do so, raw images underwent a two-step geometric correction. First, 3D real-world surface coordinates were projected onto the 2D image plane using the camera's intrinsic (K) and extrinsic ($R[t]$) calibration matrices, incorporating focal length, yaw, and pitch metadata. Second, a homographic transformation rectified the perspective, normalizing the view of the facade. Finally, to ensure data quality, a fine-tuned MobileNet CNN classified and discarded images where no building was visible due to obstructions by foreground objects (e.g., street trees or traffic) or incorrect images (extracted images of walls that are not outside exposed) (Kramm et al., 2026).

Oblique Images of wall surfaces were added to the approach to extend the analysis to surfaces not visible from the street and be able to compare results from two different data sources. They are recorded in a way that the same regions appear in several photos. Due to this overlap, a single facade typically appears in multiple images. The first step therefore involved associating specific LoD2 wall surfaces with the available aerial images and finding the best

[1] Cyclomedia API Documentation

matches. The oblique image data was provided directly by the city of Leverkusen.

Candidate images for each wall surface were identified by verifying that the wall's spatial boundaries were contained within the camera's field of view. Additionally, a directional constraint was applied to ensure the wall's orientation opposed the camera's viewing direction (e.g., associating a south-facing wall with a north-facing camera). This initial preprocessing step typically yielded between three and thirty candidate images per surface. We chose to isolate three optimal images. To do so, we converted the 3D world coordinates of the LoD2 wall vertices to 2D pixel coordinates using each camera's intrinsic (K) and extrinsic ($R[t]$) calibration matrices and determined the wall area in pixels. This was further refined by calculating the angle between the wall's surface normal and the direction of the camera's optical axis. To mitigate high-perspective distortion, a cut-off threshold of 60° was set for this angle, as 0° would face the wall straight on, and over 60° the perspective would distort the image too much. From the remaining valid set, the three images exhibiting the largest projected pixel area were selected for further analysis. Based on the calculated pixel coordinates, the selected wall surfaces were cropped from the respective oblique images. To keep the methodology aligned with the preprocessing of street view images, the raw crops were then rectified. A homographic transformation was applied to convert the angled aerial perspective into an orthogonal, front-facing view.

Despite preprocessing, non-visible walls persist in the dataset and risk skewing detection results. Consequently, an ImageNet-pretrained network is utilized to identify and discard these invalid samples, mirroring the methodology applied to street view imagery.

Table 1. Dataset overview: Comparative count of assessed walls and source imagery.

Before Preprocessing		
Number of buildings	1,342	
Number of walls	9,009	
exposed to outside	5,744	
After Preprocessing	**Street View**	**Oblique Aerial**
Number of unique walls:		
with extracted images	1,642	5,537
with valid images	923	1,522
with results	884	1,432

4.2 Factor calculation

To accurately assess a wall's suitability for greening, several factors are calculated using a combination of geometric analysis and computer vision techniques.

Geometric Analysis Many factors can be derived from the LoD2 data directly, such as the *height*, which is calculated as the difference between the highest and lowest z-value of the wall coordinates in the LoD2 data. To determine the *total surface area* of a wall, standard library limitations regarding 3D planar geometries are circumvented by employing a vector cross-product method: The wall surface is decomposed into component triangles, which are summed. By doing so, accurate area calculation is ensured even for complex shapes. Further, it is assessed whether at least two surface vertices fall within a $\pm 10cm$ tolerance buffer of the building's ground surface z-coordinates, which is called *ground boundness*. To classify a wall's *orientation* (i.e., cardinal direction North, East, South, or West), the angle between the wall's surface and true North is calculated. Since buildings in urban areas often obstruct one another, *outside exposure* determines if a wall is actually visible. This is done by measuring the distance between the particular wall and all buildings within a radius of $5m$. If any building falls into this radius, the wall is marked as not visible.

Computer Vision Analysis To analyze the composition of the facade and separate solid wall from openings, a computer vision model detects facade elements such as windows, doors or balconies within images of the facade (see Figure 3 (b)) returning bounding boxes for each element found. For street view images, a custom trained *Detectron2*[2] net is utilized and for oblique images a custom trained *YOLOv8*[3] net. If multiple views for a wall are available, bounding boxes are fused prior to the calculation of WWR and SWA (see Subsection 4.3). Following object detection and fusion, the derivation of values proceeds in pixel space. The LoD2 wall vertices are projected onto the image plane to define the facade's bounding polygon (see Figure 4 (a)). The cumulative area of all detected bounding boxes is subtracted from this projected polygon representing the SWA in pixels (see Figure 4 (b)).

(a) (b)

Figure 4. In Figure (a), the window (label 0) and door (label 1) detection in red on a rectified building wall can be seen, along with the projected surface bounding box in blue. Figure (b) depicts the resulting polygon obtained by subtracting the detected window and door surfaces from the original outside surface polygon.

The proportion of solid wall to absolute wall area is then determined by dividing the SWA pixel area (A_{SWA_px}) by the total pixel area (A_{total_px}) of the projected polygon, to be then used to calculate the Window-to-Wall Ra-

[2] https://detectron2.readthedocs.io/en/latest/

[3] https://docs.ultralytics.com/de/models/yolov8/

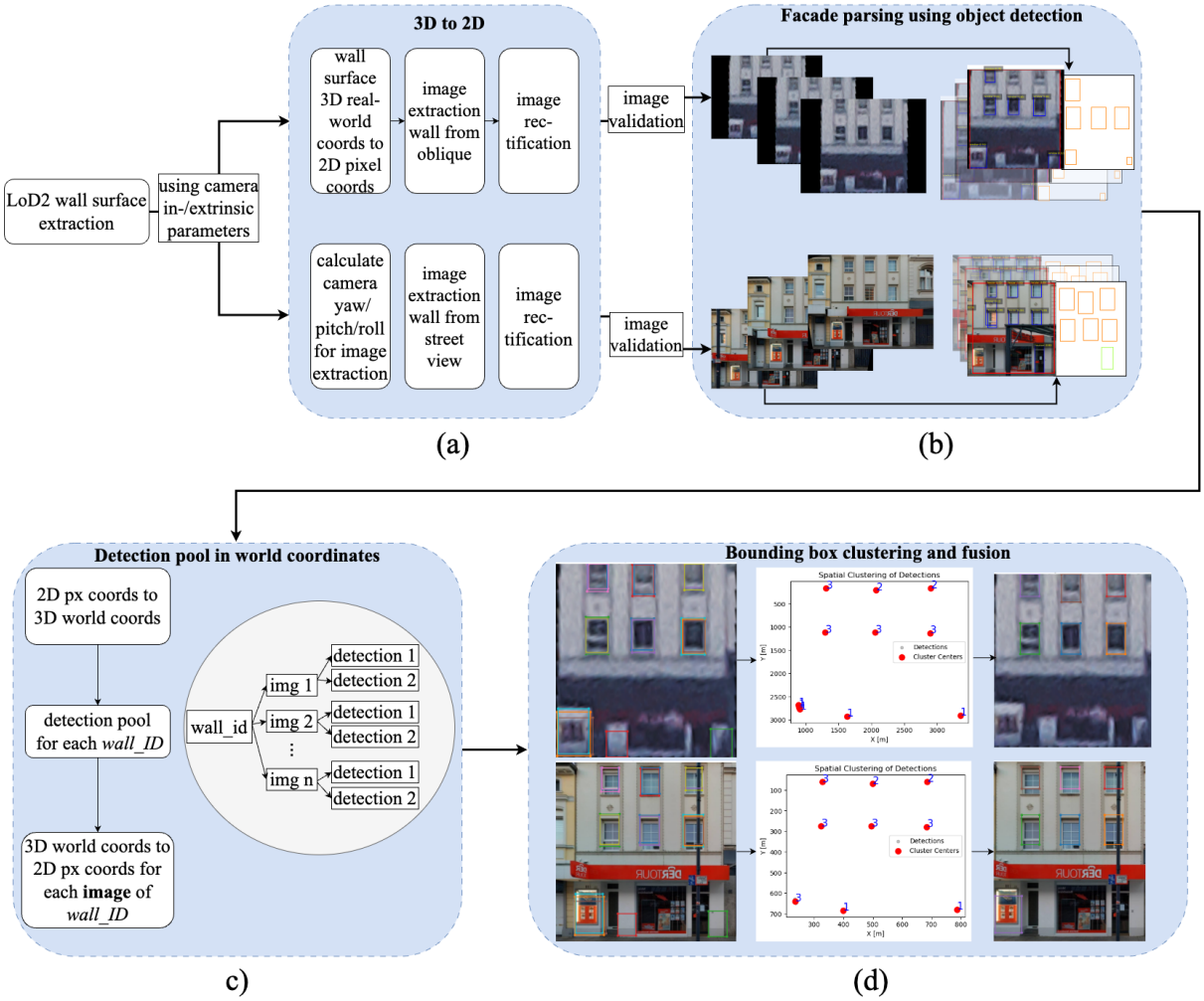


Figure 3. The pipeline for street view images (bottom) and oblique images (top). LoD2 data is used to extract wall surfaces which are then used to extract images of all visible walls. Object detection is used to parse facade elements. A multi-view detection pool is created for each wall by using 2D to 3D reprojection. For each image, the projections are recalculated separately and fused into one box for each facade element.

tio (WWR). For this study[4], the WWR is geometrically defined as the complement of this solid wall fraction (see Equation 1):

$$WWR = 1 - \frac{A_{SWA\_px}}{A_{total\_px}} \tag{1}$$

The final SWA in m^2 is obtained by multiplying the complement of the WWR by the wall's total surface area in m^2 derived from the LoD2 geometry (see Equation 2). Any wall surfaces for which valid images could not be extracted or validated are assigned NaN values for both metrics.

$$SWA_{m^2} = (1 - WWR) \times A_{total\_m^2} \tag{2}$$

[4]The definition of WWR varies by application. While energy modeling requires a distinction between glazing and window frames, this study treats the entire window (and further openings such as doors) as a void, since any aperture increases greenery maintenance complexity and therefore cost.

4.3 Multi-view fusion

Due to occlusions by trees or higher buildings, it may happen that not all facade elements can be detected from one perspective. Therefore, we perform a multi-view fusion for all detections of a wall. For walls visible from both oblique and street view perspectives, this involved the synthesis of up to six distinct views, while single-perspective walls can still benefit from the fusion of up to three images. In contrast to related studies that rely on machine learning for region proposal or object localization, we create a deterministic geometric approach exploiting the camera's calibration parameters to establish a bidirectional mapping between world and pixel coordinates. This allows for the precise calculation of pixel coordinates for all detections, bypassing the need for stochastic region-proposal networks. To fuse all detections of a wall, we first create a detection pool containing all detections in wall-aligned real-world coordinates. Pixel-space detections are projected onto the wall plane using each image's homography, en-

abling consistent fusion across images. This detection pool is then reconverted to 2D pixel space for each image separately (see Figure 3 (c)). To find and group detections of single facade objects, the centroids of all bounding box pixel values are clustered using the density-based clustering algorithm *(DBSCAN)*. All boxes of a single facade object are fused and aligned in a two-step post-processing pipeline: First, overlapping detections for the same element are aggregated by calculating the coordinate-wise median $(x_{min}, y_{min}, x_{max}, y_{max})$. We select the median over mean-based, aiming for more robustness against outliers. Second, to reflect the structural regularity of architectural facades, the resulting boxes undergo a grid alignment process. DBSCAN again analyzes the vertical and horizontal coordinates of all elements to identify dominant rows and columns. The edges of each bounding box are then snapped to the median of their respective cluster, correcting spatial jitter and enforcing a coherent rectilinear grid structure (see Figure 3 (d)). Using the fused bounding boxes, the values for WWR and SWA are calculated for each image as described in Subsection 4.2. These values are then averaged first within each data source yielding one result per source, and subsequently these are averaged for the final result for each wall.

4.4 Vertical Greenery Potential Index

To prioritize facades most suitable for greening, prior work established a potential index. This composite metric aggregates three key variables: Solid Wall Area, Window-to-Wall Ratio, and orientation. This scoring system identifies facades where vegetation would enhance evaporative cooling and shading, thus offering the most significant contribution to mitigating urban heat stress.

To obtain a broader spectrum of index values, power transformations are applied prior to normalization. SWA is adjusted with an exponent of 0.3 followed by min-max scaling. WWR is similarly transformed with an exponent of 0.5. As a high WWR should result in a lower potential index, the complement $(1 - WWR)$ is utilized for the final calculation. Orientation was scored discretely based on solar exposure potential, by assigning a fixed value to each orientation prioritizing south-facing walls (0.5) over east/west (0.35) and north (0.15). The final index is computed as a weighted linear combination:

$$Index_{pot} = 0.5 * SWA_{\text{norm}} + 0.4 * WWR_{\text{inv\_norm}} + 0.1 * Orientation \quad (3)$$

5 Evaluation and Results

A primary constraint of this work is the absence of a ground-truth dataset for the building stock in Leverkusen (e.g., a LoD3 model containing facade openings). To be able to assess the accuracy of the proposed pipeline, a ground-truth dataset was established with a sample of 50 facades containing both street view and oblique images. For these, all visible facade elements (e.g., windows, doors) were manually annotated within the projected LoD2 surface. These annotations were then processed using the calculation pipeline described in Section 4, yielding ground-truth values for WWR and SWA. Complementing this, a comparative analysis between the two data sources (street view and oblique imagery) was conducted. By evaluating the cross-modal consistency between street view and oblique aerial imagery, we aim to determine whether they yield comparable results or exhibit significant differences. Quantifying this inter-source variability is critical for establishing the conditions under which results yielded by a single-source can be considered trustworthy for city-wide assessments, particularly where dual-perspective coverage for walls is unavailable.

5.1 Geometric Misalignment

The estimation of WWR and SWA relies on the projection of LoD2 wall surface geometries onto 2D image planes. This introduces a significant source of uncertainty, as the projection often deviates from the actual building facade. This can manifest in the projected geometries exceeding the actual facade boundaries or polygons suffering from positional offsets, where the projected area is approximately accurate but fails to align with the visual edges of the building. (see Figure 5 (b) and (c)). Such misalignment can lead to the inclusion of background noise or the overestimation of wall areas. According to the city's geodata office, approximately 30% of the LoD2 dataset is affected by such spatial imprecision. To mitigate these effects, the projection should be corrected first. A possible solution could be to use the initial LoD2 projection as a Region of Interest (ROI) for facade segmentation. This refinement step would allow the facade boundaries to snap to the actual visual edges, ensuring higher accuracy in the subsequent factor calculation.

5.2 Dataset composition and Image Availability

From the initial dataset comprising approximately 9,000 wall surfaces, our analysis identified 5,700 surfaces with outside exposure, meaning the remaining walls are directly adjacent to neighboring buildings (see Table 1). Data acquisition rates varied significantly by source: The street view perspective yielded images for only ≈30% of the exposed surfaces. In contrast, oblique imagery achieved near-universal coverage, substantiating the study's primary motivation to achieve greater coverage of all walls using oblique images.

Following acquisition, a CNN validated if an image displays a house or a wall and discarded images that did not. Post-validation, the viable dataset for oblique imagery exhibited a significant decline, with only 1,500 walls remaining in the dataset. This substantial reduction

Figure 5. Detection comparison for street view (above) and oblique (below) images. The blue bounding boxes indicate detections from the object detection network. It becomes apparent that not all facade elements are always recognized in an image. The red bounding box indicates the inaccurate, projected wall surface from the LoD2 data.

stems from the oblique extraction pipeline's inherent sensitivity to the segmented geometry of the LoD2 models. Street view image acquisition relies on calculating camera extrinsics (yaw and pitch) to find eligible recording points that center the target, retrieving the full sensor's field of view. Consequently, even small wall segments are captured within a larger, recognizable building context, satisfying the CNN's validation criteria. Conversely, the oblique extraction pipeline utilizes a strict 3D-to-2D vertex projection to isolate the specific wall surface. Due to the fragmented nature of LoD2 modeling, where buildings are often composed of multiple parts, this strict cropping frequently yields low-resolution snippets lacking sufficient semantic features for validation (see Figure 6). This issue highlights a critical dependency on the quality of GML modeling, as geometric fragmentation persists despite preprocessing efforts to merge coplanar wall segments.

Figure 6. Example cut outs from oblique imagery for fragmentedly modelled wall surfaces.

5.3 Source-Specific Limitations and Variance

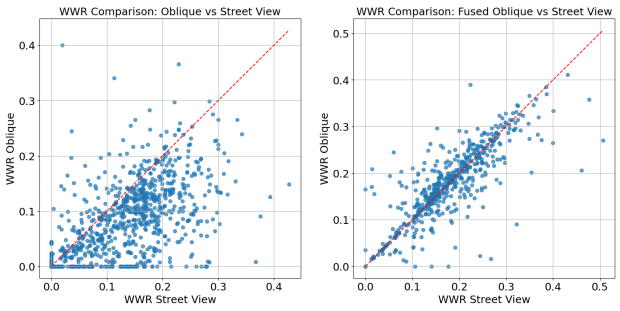
Before comparing across perspectives, we assessed the internal consistency of the derived SWA and WWR within each image type. For both WWR and SWA, the typical per-wall variation across multiple images was low with a standard deviation of about 2% for oblique and street view images. This indicates consistency within images of the same perspective. For a small number of walls, both perspectives nevertheless included a small number of walls with substantially higher variability, likely due to partial occlusion, complex geometry, or limited coverage in some images. While susceptibility to occlusion is a shared limitation, the nature of the obstruction differs by source: Street view data is frequently obstructed by objects such as vegetation or vehicles, while oblique imagery suffers from occlusions caused by adjacent buildings and tree canopies. These obstructions can negatively impact the calculation of the SWA, as facade objects are often hidden and therefore not detected by the object detection. Even though this limitation concerns both data sources, our analysis shows that if dual coverage is available for a specific wall, occlusions do not necessarily occur in both perspectives, reinforcing the benefit of using multiple data sources.

Table 2. Differences in wall-level estimates between approaches

	Individual: SV vs. Obl.	**Fused: SV vs. Obl.**
Median Diff. WWR in %	-4.3	**0**
Max Diff. WWR in %	22	19.4
Median Diff. SWA in m^2	3.4	**-0.1**
Max Diff. SWA in m^2	88.3	85.3

5.4 Comparative Analysis: Street View vs. Oblique Imagery

To evaluate the proposed fusion pipeline and compare results from both data sources, a cross-modal validation was performed by comparing the SWA and WWR estimations derived from isolated walls in both street view and oblique aerial imagery. A comparison of factors derived independently from oblique and street view imagery reveals systematic differences at the wall level. For WWR, oblique estimates are consistently lower than those obtained from street view, indicating a negative bias (see Figure 7 (a)). Consequently, the SWA exhibits a similar pattern, with oblique estimates exceeding street view (see Figure 8 (a)). While the magnitude of disagreement varies between walls, median differences indicate small discrepancies between the two perspectives, and in individual cases, large outliers can be observed. These findings highlight the sensitivity of estimates to different viewpoints (see Table 2).



(a) (b)

Figure 7. Comparing results for the window-to-wall ratio (WWR) before (a) and after (b) fusion. Results after fusion show less bias between both image sources.

(a) (b)

Figure 8. Comparing results for the Solid Wall Area before (a) and after (b) fusion cut off at 150 for readability. Results after fusion show less bias between both image sources.

However, the practical impact of these discrepancies on the assessment of greening potential is less severe than the raw variance suggests, as the independent SWA estimates demonstrate general agreement even prior to fusion (see Figure 8 (a)). Since the SWA is derived from the total wall area multiplied by the complement of the WWR (see Equation 2), a deviation in the ratio (e.g., Δ=0.1) often translates to a manageable difference in absolute square meterage. This indicates that, despite the identified bias, both data sources independently yield sufficiently accurate approximations to serve as valid baselines for estimating vertical greenery potential, although fusion also reinforces the consistency in this case (see Figure 8 (b)).

The systematic underestimation of WWR in aerial images can be linked to a lower density of object detections within the same wall compared to the street view perspective (see Figure 9). There are two explanations for this: first, the coarser spatial resolution of oblique aerial imagery can obscure small features. Second, the acute viewing angles of the aerial cameras towards wall surfaces can lead to roof eaves or structural overhangs covering the upper parts of the facade (see Figure 5 (c)). Furthermore, the underlying neural networks are subject to classification errors where misclassifications or missed detections in either modality may introduce a marginal degree of noise into the final results.

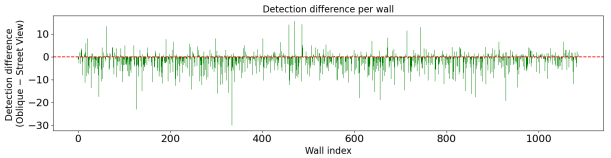


Figure 9. Comparing the number of detections in street view and oblique images. 0 indicates the same number of detections in street view and oblique images, negative numbers indicate more detections in street view images.

To mitigate the noise and enhance the methodological robustness, a multi-view fusion strategy was implemented by aggregating and aligning detections from all available images per wall. The fusion approach substantially improves agreement between the two data sources (see Figure 7 (b) and 8 (b)). With the median differences reduced to near zero for both WWR and SWA, the systematic bias becomes negligible, resulting in nearly unbiased and highly consistent estimates across most facades. The improved agreement is achieved by a more consistent number of detections across all images of a wall, demonstrating the increased robustness of multi-view integration against viewpoint-dependent errors.

Nevertheless, the applied fusion methodology in this study is not without limitations. First, the aggregation logic remains vulnerable to detection artifacts: if the input data contains a high frequency of imprecise bounding boxes, a common issue in oblique imagery where detections often fail to snap precisely to window edges due to the low resolution, the fusion algorithm inadvertently reinforces these geometric inaccuracies rather than filtering them out. Second, the clustering mechanism is highly sensitive to the selected parameters, specifically the spatial distance threshold (ϵ) of the DBSCAN algorithm, where suboptimal tuning can lead to either the over-segmentation of single elements or the erroneous merging of adjacent windows. Finally, while Region-of-Interest (ROI) refinement strategies

used in related research could theoretically mitigate such alignment errors, their practical utility remains uncertain in this context. Since the primary impairment in oblique detections likely stems from inherent low resolution, geometric post-processing alone may be insufficient to better align the bounding boxes.

5.5 Ground-truth comparison

Finally, the results of the fusion pipeline are tested against a manually annotated ground-truth dataset of 50 facades. This set excludes facades with severe occlusions, however, partially obscured elements were annotated to their full inferred extent. Furthermore, the projected LoD2 surface geometry was utilized to derive ground-truth WWR and SWA despite known inaccuracies. Thus, this validation isolates the performance of the detection and fusion algorithms, validating internal consistency rather than absolute conformity to physical reality. The residuals were calculated by subtracting the calculated value from the ground-truth value. The WWR analysis reveals a high degree of accuracy, with the majority of residuals falling within the narrow range of ± 0.1 (see Figure 10 (a)). Translating these ratios onto the physical facade, resulting deviations lie between $\pm 10 m^2$ (see Figure 10 (b) and (c)). Given the scale of urban planning, this margin of error is well within acceptable limits for identifying high-potential surfaces.

5.6 Index Calculation

The potential index was calculated for a total of 1,889 walls (see Figure 11). Where available, values derived from the multi-perspective fusion were prioritized; otherwise, the value from the single-source calculation was used.

A qualitative review of the indexed walls revealed that results derived from street view imagery consistently yielded surfaces that suggest to be ideal candidates for greening (see Figure 12 (top row)). These candidates are characterized by a lack of windows and large surface areas. Nevertheless, the majority of results from walls ranked highly in the index are derived from oblique images. Looking at these walls, they exhibit detection omissions or occlusions by vegetation, which emerged as the leading source of error, along with the method's susceptibility to producing small or occluded image cut-outs (see Figure 12 (bottom row)). This indicates that these top-ranked aerial candidates may not be as suitable for vertical greenery as the index initially suggests. However, this does not generally question the calculation of the index, but emphasizes the importance of thorough data cleaning for the calculation and highlights the importance of object recognition performance for facade segmentation and image validation.

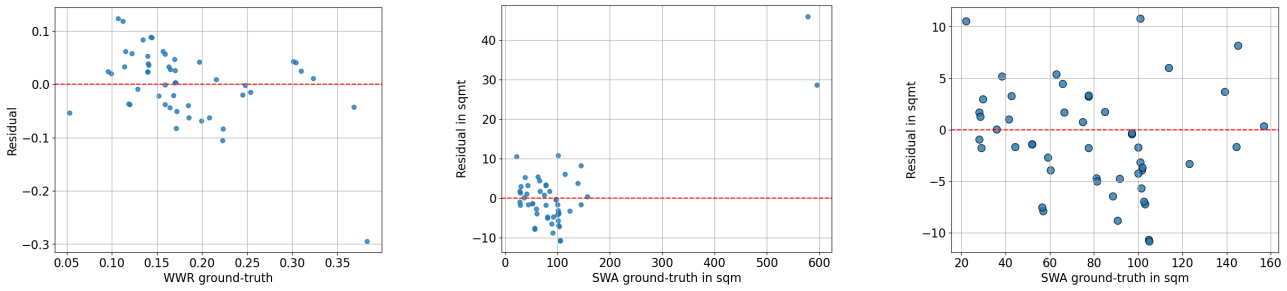
6 Discussion and Outlook

Our findings indicate a high level of agreement between the individual street view and oblique perspectives. By synthesizing detections from multiple angles, the approach significantly increases robustness against viewpoint-specific occlusions, leading to less systematic bias in WWR and SWA estimates. Crucially, this fusion process operates on multiple levels: while dual-source coverage allows for the integration of up to six images per wall, the pipeline still successfully fuses up to three images when only a single data source is available. While the first provides the highest degree of occlusion mitigation, its application is limited to walls with dual coverage. Nevertheless, the method's reliability is strongly validated by the ground truth analysis, which shows minimal deviation from manual annotations.

The evaluation of the vertical greenery potential index further highlights source-specific differences. While estimates of facades within the top index rankings derived from SVI yield suitable candidate walls for greening, the analysis of walls, whose values originate from oblique images, has shown that occlusions and omitted detections may yield misleadingly high suitability scores and require further validation. Yet the high occurrences of aerial-derived facades within the top index rankings stresses the superior spatial coverage.

While the proposed computational framework is transferable to any region with adequate data availability, the specific suitability parameters of this study were formulated and applied for central European cities. Here, the prioritizing of south-facing facades optimizes cooling effects and plant vitality, whereas in southern Europe, high solar irradiance can become a stressor for plants. In such regions, the weighting logic would need to be adjusted to favor more shaded orientations. Similarly, the typological context of this study was low-rise building blocks. To find suitable walls in high-rise environments, further variables such as wind loads must be incorporated.

At last, the inherent inaccuracies in the LoD2 modeling introduce errors in both image extraction and factor calculation. As the LoD2 vertices are used to project wall surfaces onto oblique images for extraction, the granular partitioning of buildings in the GML model often results in tightly cropped, low-resolution image patches that lack sufficient semantic features for validation. The analysis also revealed a lower object detection rate and poorer bounding box quality in the oblique dataset compared to street view. This performance gap is likely attributable to a combination of lower spatial resolution and the fact that current computer vision architectures are predominantly pre-trained on higher resolution datasets. Furthermore, the current fusion logic remains sensitive to the quality of facade object detections, as imprecise bounding boxes can propagate errors rather than resolving them.



(a) (b) (c)

Figure 10. Quantitative comparison of the fused pipeline against manually annotated ground truth. Residual plots demonstrate minimal deviation in WWR across the dataset, with the exception of a single outlier. Consequently, the discrepancies in absolute Solid Wall Area (m^2) remain negligible. Plot (c) is a zoom-in version of plot (b) two without outliers.

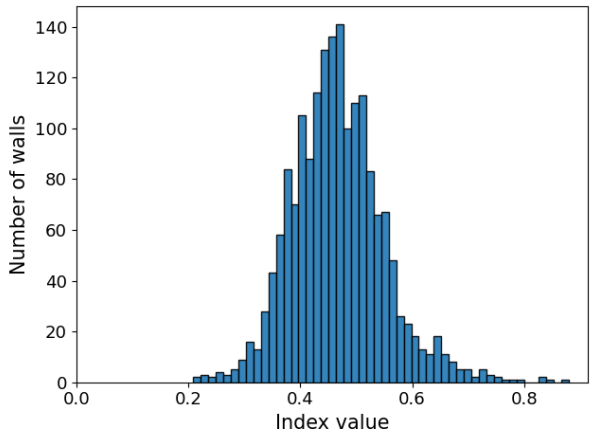


Figure 11. The calculated index for all walls in the test area.

Figure 12. Example for walls within the top ten of the index. Candidates whose results are derived from street-view imagery (top row) are identified as promising targets, whereas candidates whose results are derived from oblique imagery (bottom row) are predominantly affected by occlusions, such as vegetation or roof eaves.

Future research should therefore decouple the extraction pipeline from geometric dependencies and find the true boundaries of the facade by employing semantic segmentation, where the raw LoD2 projection could still serve as initialization. To enhance the reliability of the oblique pipeline, the geometric merging of LoD2 wall segments during pre-processing should be refined to eliminate fragmented or erroneous cutouts. Providing the facade detection and validation network with more coherent facade regions could improve filtering accuracy and overall detection performance. If advanced preprocessing can successfully refine wall geometries and thus mitigate extraction noise, oblique imagery could potentially serve as a sufficient standalone data source in urban areas where dual-data coverage is not feasible. Despite these challenges, the fusion of terrestrial and aerial data represents a vital step toward a comprehensive digital inventory of urban vertical greenery potential.

7 Data and Software Availability

Data. The data used in this work was mostly provided by the city of Leverkusen and is not ours to share due to licensing restrictions. For inquiries regarding oblique images, refer to *Sachgebiet Geodatenmanagement Leverkusen* and regarding street view *Cyclomedia company*.

The Level of Detail 2 city model is openly available and was downloaded via https://www.opengeodata.nrw.de/produkte/geobasis/3dg/lod2_gml/lod2_gml/. An example file can be found in the repository.

Code. The code for the approaches described in this paper can only be made available in limited form for the following reasons:

- Proprietary API and Data: Certain components of the pipeline are derived from proprietary data and APIs that are not publicly accessible. Releasing the source code without appropriate authorization or without redacting these elements could lead to contract violations.

- API Keys: To obtain street view images from the *street smart* API, confidential credentials are needed. Making the code public would require removing or masking these elements, which would limit the application's functionality.

- Licensing restrictions: The trained models cannot be made available due to licensing restrictions. Since they were trained on proprietary data, their owner did not allow us to publish the models.

The code that can be made available without restrictions can be found via https://github.com/aruscha-k/Quantify-VGS and contains the following items:

- Example LoD2 data downloaded from the above link
- LoD2 preprocessing and derivation of factors from LoD2 data
- Fusion of street view and oblique detections
- Computation for the potential index

8 Declaration of Generative AI in Writing

The authors declare that they have used Generative AI tools in the preparation of this manuscript for language editing and sentence structure. All intellectual and creative work, including the analysis and interpretation of data, is original and has been conducted by the authors without AI assistance.

9 Acknowledgements

This paper was created as part of the "Connected Urban Twins" (CUT) project, funded by the German Federal Ministry of the Interior, Building and Community.

References

Biljecki, F. et al.: The variants of an LOD of a 3D building model and their influence on spatial analyses, ISPRS Journal of Photogrammetry and Remote Sensing, 116, 42–54, https://doi.org/10.1016/j.isprsjprs.2016.03.003, visited on 12/30/2024, 2016.

BuGG Bundesverband GebäudeGrün e. V.: Planungshinweise für Fassadenbegrünung, https://www.gebaeudegruen.info/gruen/fassadenbegruenung/planungshinweise, visited on 10/10/2024, 2024.

Bustami, R. A. et al.: Vertical greenery systems: A systematic review of research trends, Building and Environment, 146, 226–237, https://doi.org/10.1016/j.buildenv.2018.09.045, visited on 09/23/2024, 2018.

Gaw, L. Y., Chen, S., Chow, Y. S., Lee, K., and Biljecki, F.: COMPARING STREET VIEW IMAGERY AND AERIAL PERSPECTIVES IN THE BUILT ENVIRONMENT, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W3-2022, 49–56, https://doi.org/10.5194/isprs-annals-X-4-W3-2022-49-2022, 2022.

Heaviside, C., Macintyre, H., and Vardoulakis, S.: The Urban Heat Island: Implications for Health in a Changing Environment, Current Environmental Health Reports, 4, 296–305, https://doi.org/10.1007/s40572-017-0150-3, 2017.

Kramm, A., Friske, J., and Peukert, E.: Detecting Floors in Residential Buildings, in: KI 2023: Advances in Artificial Intelligence, edited by Seipel, D. and Steen, A., vol. 14236 of *Lecture Notes in Computer Science*, pp. 130–143, Springer Nature Switzerland, Cham, https://doi.org/10.1007/978-3-031-42608-7_11, visited on 03/04/2025, 2023.

Kramm, A. et al.: Automated Estimation of Urban Vertical Greenery Potential, https://doi.org/10.21203/rs.3.rs-8548920/v1, preprint (Version 1) available at Research Square, 2026.

Krylov, V. A. and Dahyot, R.: Object Geolocation Using MRF Based Multi-Sensor Fusion, in: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 2745–2749, IEEE, Athens, https://doi.org/10.1109/ICIP.2018.8451458, 2018.

Köhler, M.: Green facades—a view back and some visions, Urban Ecosystems, 11, 423–436, https://doi.org/10.1007/s11252-008-0063-x, visited on 09/24/2024, 2008.

Liu, H., Li, W., and Zhu, J.: Translational Symmetry-Aware Facade Parsing for 3-D Building Reconstruction, IEEE MultiMedia, 29, 38–47, https://doi.org/10.1109/MMUL.2022.3195990, visited on 12/03/2024, 2022.

Liu, H. et al.: DeepFacade: A Deep Learning Approach to Facade Parsing With Symmetric Loss, IEEE Transactions on Multimedia, 22, 3153–3165, https://doi.org/10.1109/TMM.2020.2971431, visited on 12/03/2024, 2020.

Lu, Y. et al.: A deep learning method for building façade parsing utilizing improved SOLOv2 instance segmentation, Energy and Buildings, 295, 113 275, https://doi.org/10.1016/j.enbuild.2023.113275, visited on 12/03/2024, 2023.

Ma, W. et al.: Pyramid ALKNet for Semantic Parsing of Building Facade Image, IEEE Geoscience and Remote Sensing Letters, 18, 1009–1013, https://doi.org/10.1109/LGRS.2020.2993451, visited on 12/04/2024, 2021.

Nassar, A. S., D'Aronco, S., Lefèvre, S., and Wegner, J. D.: GeoGraph: Graph-Based Multi-view Object Detection with Geometric Cues End-to-End, in: Computer Vision – ECCV 2020, edited by Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., vol. 12352, pp. 488–504, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-58571-6_29, series Title: Lecture Notes in Computer Science, 2020.

Oliveira, J. A. P. D. et al.: Innovations in Urban Green and Blue Infrastructure: Tackling local and global challenges in cities, Journal of Cleaner Production, 362, 132 355, https://doi.org/10.1016/j.jclepro.2022.132355, visited on 01/21/2025, 2022.

Perez, G., Coma, J., Martorell, I., and Cabeza, L. F.: Vertical Greenery Systems (VGS) for energy saving in buildings: A review, Renewable and Sustainable Energy Reviews, 39, 139–165, https://doi.org/10.1016/j.rser.2014.07.055, 2014.

Safikhani, T. et al.: Thermal Impacts of Vertical Greenery Systems, Environmental and Climate Technologies, 14, 5–11, https://doi.org/10.1515/rtuect-2014-0007, visited on 09/24/2024, 2014.

Santamouris, M.: Cooling the Cities – A Review of Reflective and Green Roof Mitigation Technologies to Fight Heat Island and Improve Comfort in Urban Environments, Solar Energy, 103, 682–703, https://doi.org/10.1016/j.solener.2012.07.003, 2014.

Sezen, G. et al.: Deep Learning-Based Door and Window Detection from Building Façade, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B4-2022, pp. 315–320, https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-315-2022, visited on 12/03/2024, 2022.

Sun, Y. et al.: DeepWindows: Windows Instance Segmentation through an Improved Mask R-CNN Using Spatial Attention and Relation Modules, ISPRS International Journal of Geo-Information, 11, 162, https://doi.org/10.3390/ijgi11030162, visited on 12/03/2024, 2022.

Wegner, J. D., Branson, S., Hall, D., Schindler, K., and Perona, P.: Cataloging Public Objects Using Aerial and Street-Level Images — Urban Trees, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6014–6023, https://doi.org/10.1109/CVPR.2016.647, iSSN: 1063-6919, 2016.

Xu, H. et al.: LOD2 for energy simulation (LOD2ES) for CityGML: A novel level of details model for IFC-based building features extraction and energy simulation, Journal of Building Engineering, 78, 107 715, https://doi.org/10.1016/j.jobe.2023.107715, visited on 12/29/2024, 2023.