Validating a window view simulation engine based on open data and open source using semantic segmentation

Anna-Maria Bolte [1, *], Thomas Kistemann [2, 3], Theo Kötter [1], Youness Dehbi [4]

[1] Institute of Geodesy and Geoinformation, University of Bonn, Germany – (ambolte, tkoetter)@uni-bonn.de
[2] Institute for Hygiene & Public Health, GeoHealth, University Hospital Bonn, Germany – Thomas.Kistemann@ukbonn.de
[3] Department of Geography, University of Bonn, Germany
[4] Computational Methods Lab, HafenCity University Hamburg, Germany – youness.dehbi@hcu-hamburg.de
[*] Correspondence: ambolte@uni-bonn.de

Keywords: GWVI, window views, urban green spaces, urban morphology, semantic segmentation, Cityscapes.

Abstract:

The visual access to urban green spaces through window views plays a key role in increasing well-being, particularly for those with limited mobility. This study verifies a window view simulation engine around the Green Window View Index (GWVI) that combines open source approaches with open geospatial data. Using a pretrained DeepLab V3+ model on Cityscapes data set for semantic segmentation, the study compares the accuracy of simulated window views to photorealistic semantic segmentations. A total of 40 window views were examined, with 0.1 m and 2.0 m distance to the window. The validation metrics consist of overall accuracy (OAcc), mean accuracy (mAcc), mean intersection over union (mIoU), and individual IoU values for vegetation, sky, and buildings. The statistics show an mIoU of 0.53, with class-specific IoU values of 0.52 for vegetation, 0.64 for the sky, and 0.43 for buildings, an OAcc of 0.68, and an mAcc of 0.74. The approach has a low variance in visibility values, with a minor underestimating of vegetation (-6%), and an overestimation of sky (+5%) and buildings (+3%). These findings indicate that the simulation engine performs well, outlining its potential for analyzing window views in a variety of urban scenarios. Future large-scale crowdsourcing experiments are recommended to statistically support these findings.

1. Introduction

Window views are often the only way to access urban green spaces for vulnerable populations with a limited radius of movement (Pijpers and van Melik, 2020). In light of the spatial consequences of increasing climate disasters, such as heat waves, or pandemics, including Covid-19, this access alternative is becoming increasingly important for sustainable and resilient open space planning (Amerio et al., 2020; Basu et al., 2024; Haaland and Konijnendijk van den Bosch, 2015). The positive effects of green window views on urban dwellers regarding health, cognitive abilities, and social well-being have been demonstrated in numerous studies. In addition, empirical evidence has shown that these views can enhance life satisfaction and increase property values (Bolte et al., 2025; Meng and Wang, 2024). Existing methodologies for automated quantification and analysis of window views employ machine learning image segmentation techniques grounded in advanced photorealistic city information models (CIM) or big data (Swietek and Zumwald, 2023; Li et al., 2022; 2024; Peng et al., 2025). These data bases facilitate a robust approach for evaluating window views; however, they are primarily available for urban areas characterized by high density and high-rise buildings (Li et al., 2022; 2024; Peng et al., 2025).

In order to capture a range of urban morphologies with diverse densities and degrees of sealing, the approach of the Green Window View Index (Bolte et al., 2024a) uses official digital geospatial data, such as the official German cadastral information system (ALKIS), semantically segmented 3D city models CityGML, LiDAR point clouds, and Sentinel-2 landcover data, which are continuously transmitted and harmonized by the public sector for national, regional, and municipal planning processes (BKG, 2024). The three-steps visibility analysis has been applied in several municipal case studies. The analysis of numerous million window views from various urban structure types provided conclusions for urban green space planning in residential environments (Bolte et al., 2024b; 2025).

The aim of this paper is to verify the reliability of the window view simulation engine to ensure the validity of the previous analyses. For this purpose, a comparison of simulated window views with semantically segmented photorealistic window views as ground truth is performed to address the following questions:

1. How high is the overall accuracy of the simulated window views?
2. How high is the mean accuracy of visible classes in the simulated window views?
3. How high is the intersection over union of individual visible classes in the simulated window views?
4. How high is the mean intersection over union of all visible classes in the simulated window views?
5. How high is the visibility of individual visible classes in the simulated window views?

First, the present article defines the validation workflow, including the window view simulation, the creation of the photorealistic ground truth data, and the definitions of the validation metrics. The results are subsequently presented regarding the quality of the intersection and visibility values. The validation results are then interpreted with existing CIM and machine learning methodologies and are critically discussed. Finally, a conclusion of the results is presented.

2. Methodology

The validation workflow is divided into three steps (refer to Figure 1). The first step involves simulating individual window views using the simulation engine. The second step involves creating semantically segmented photorealistic ground truth data and subsequently evaluating the window views using validation metrics.

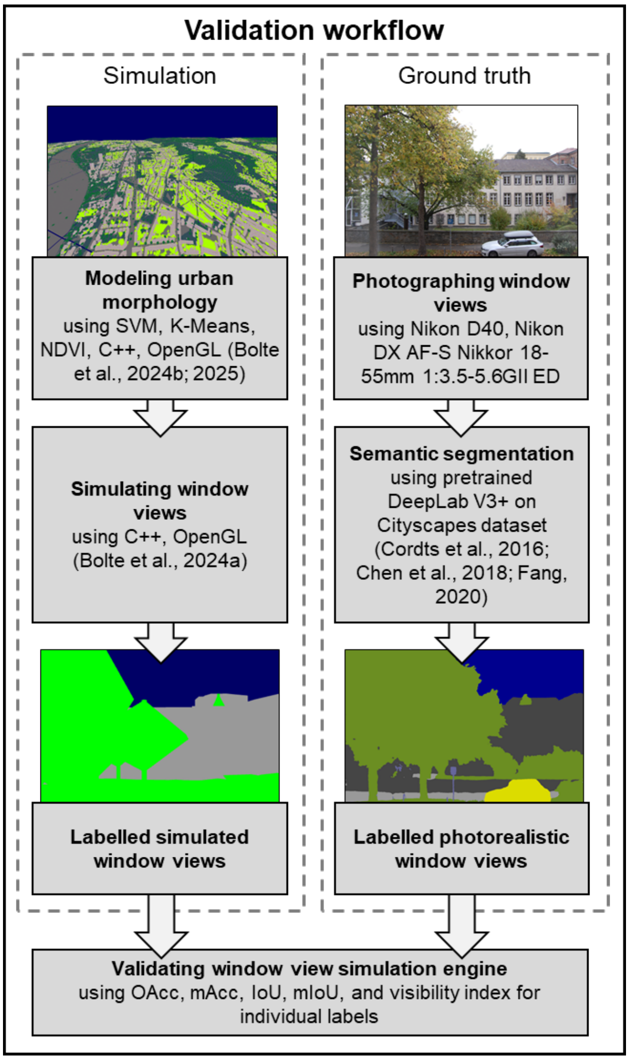

Figure 1. Validation workflow of window view simulation engine, the example window view shown on the right sight was taken at a distance of 0.1 m to the window.

2.1 Simulating window views

In order to generate simulated window views including visible vegetation, sky, and buildings, the initial step involves modeling the built-up and greened urban morphology, including the following elements: topography, property boundaries, buildings, as well as flat and tall vegetation. The official German cadastral information system (ALKIS), segmented 3D city models (CityGML) with level of detail (LoD) 2, LiDAR point clouds, and Sentinel-2 landcover data were applied in this process. Tall vegetation was derived by a classification and clustering process using Support-Vector-Machine (SVM) and K-Means, flat vegetation was modelled by using Normalized Difference Vegetation Index (NDVI). Q-GIS 3.34.3 and CloudCompare v2.13.beta were used for this process. (Bolte et al., 2024b; 2025)

The present case study was limited to the street front of an institute building in the University of Bonn, Germany. The building was constructed between 1906 and 1907 and is characterized by features typical of the Neo-Renaissance style. However, the CityGML dataset with LoD 2 lacks sufficient information regarding window positioning and dimension. To address this limitation, we used an existing exterior surveying of the building, which was initially conducted for the purpose of window modernization. This approach allowed us to consider the position and dimensions of five single-sash box-type windows on the basement floor, six three-sash box-type windows on the first floor, and six double single-sash box-type windows on the second floor (refer to Figure 2).

Figure 2. Street front of investigated building including 17 considered box-type windows in the basement, first, and second floor.

The window view simulation is based on OpenGL, is developed in C++ and included the following parameters for a distance of 0.1 m to the window: a vertical field of view of 100°, an aspect ratio of 1.45, and a maximum view distance of 5,000 m (Bolte et al., 2024b; 2025). At a distance of 2.0 m to the window, parameters in Table 1 were implemented based on the exterior surveying of the building.

Parameters	Basement floor	1st floor	2nd floor
Window size	1.4 m x 1.58 m	1.88 m x 2.44 m	1.84 m x 2.44 m
Vertical field of view	43.1°	62.8°	62.8°
Aspect ratio	0.89	0.77	0.75
Max. view distance	5,000 m	5,000 m	5,000 m

Table 1. Simulation parameters at a window distance of 2.0 m.

A total of 40 window views (17 views at 2.0 m distance to the window and 23 views at 0.1 m distance to the window were simulated. The labels in the window simulations were colored as follows: vegetation (green), sky (blue), buildings (dark grey), and streets (light grey). The simulation was carried out on a 64-bit operating system, WLS Ubuntu 20.04.6 LTS WLS on Windows 11 Pro including a 12th generation Intel® Core™ i9-12900K CPU/3.20 GHz, a NVIDIA RTX A4500 graphics card, and a 128 GB SSD. The average running time was 0.05 s per window. (Bolte et al., 2024a)

2.2 Segmenting photorealistic window views as ground truth

Ground truth data was taken by using a camera equipped with an SLR lens, which was mounted on a tripod (refer to Figure 1 for technical details). The camera was directed at the center of the windows, at a distance of 0.1 m and 2.0 m from the window surface.

The images were semantically segmented using a pretrained DeepLab V3+ model on the Cityscapes dataset including the following technical details: backbone = resnet 101, batch size = 16, FLOPs = N/A, train/val OS = 16/16. The mIoU was 0.762. Python ver. 3.11 including pytorch 0.3.4 was utilized for this process. Defined labels were based on Cityscapes and were summarized as described in Table 2 (Cordts et al., 2016; Chen et al., 2018; Fang, 2020).

Considered labels	Cityscapes labels	Colors
Streets	Road, sideroad	Light grey
Buildings	Building, wall, fence	Dark grey
Objects	Pole, traffic light, traffic sign	Grey blue
Vegetation	Vegetation, terrain	Green
Sky	Sky	Blue
Humans	Person, rider	Red
Vehicles	Car, truck, bus, train, motorcycle, bicycle	Yellow

Table 2. Defined labels for semantic segmentation process.

As shown by Li et al. (2024) and Peng et al. (2025), the use of DeepLab V3+ based on outdoor Cityscapes datasets can result in segmentation inaccuracies of indoor window views. In particular, we observed incorrect segmentation of window frames, which would have resulted in the falsification of the validation of the entire window views. To address this issue, a mask was manually created for each window type to represent the corresponding window frame. These masks were then overlaid on the simulation and segmentation results, focusing the validation on the labeled window view content for views at a window distance of 2.0 m. GIMP 2.10.32 (Revision 1) was used in this process (refer to Figure 4, masked simulation and segmentation).

2.3 Validation metrics

The validation was finally conducted by assessing the overall accuracy (OAcc), mean accuracy (mAcc), mean intersection over union (mIoU), and individual intersection over union (IoU) values to visible vegetation, sky, and building. Furthermore, the proportion of visible vegetation, sky, and building within the window view for 0.1 m and 2.0 m distance to the window was quantified and compared based on the conducted simulation and the photorealistic image segmentation results (refer to Table 3, Bolte et al., 2024a; Li et al., 2022; 2024).

Equations	Parameters
$OAcc = \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i + FN_i)}$	Overall Accuracy: TP_i = true positive for class i FP_i = false positive for class i FN_i = false negative for class i
$mAcc = \frac{1}{n}\sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i + FN_i}$	Mean Accuracy: n = number of considered classes
$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}$	Intersection over Union: TP_i = true positive for class i FP_i = false positive for class i FN_i = false negative for class i
$mIoU_i = \frac{1}{n}\sum_{i=1}^{n} IoU_i$	Mean Intersection over Union: IoU_i = IoU for class i
$WVI_{id} = \frac{Visible\ Target\ Area\ in\ Window_{id}}{Area(Window_i)} * 100$	Window View Index: i = ith window d = distance to window

Table 3. Metrics considered for validation process.

3. Results

3.1 Visual comparison of window view simulation and photorealistic semantic window view segmentation

While a preliminary visual assessment indicates a promising simulation outcome, a more detailed examination uncovers discrepancies that explain the results of the metric validation (refer to subsections 3.2 and 3.3).

Primarily, the segmentation results show a more detailed representation and differentiation of the labels. For instance, the segmented window views differentiate vehicles, objects, and humans, in addition to vegetation, buildings, streets, and the sky. However, these labels are segmented incorrectly in certain areas (refer to Figure 3, basement floor, segmentation and Figure 4, basement floor, masked segmentation). Furthermore, additional segmentation errors are visible in the label “sky” (refer to Figure 3, second floor, segmentation and Figure 4, first and second floor, masked segmentation, respectively).

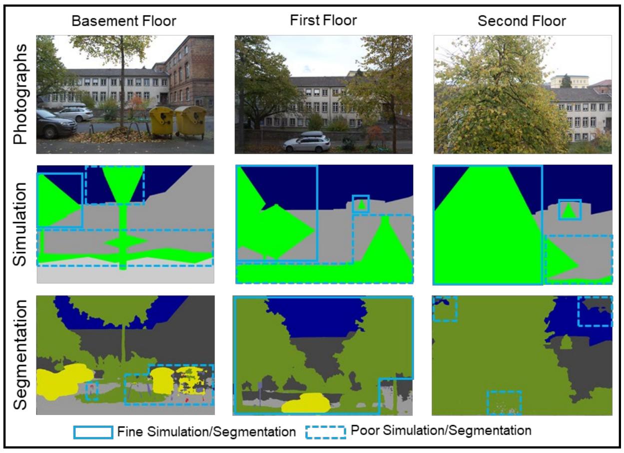

Figure 3. Exemplary comparison of window view simulation to photorealistic semantic segmentation of Deeplab V3+ pretrained on Cityscapes at a window distance of 0.1 m.

In comparison, the simulation results demonstrate a reduction in the visual complexity of building structures, such as walls or fences, due to the generalizations inherent in the modeling of urban morphology (refer to subsection 2.1). Additionally, an incomplete rendering of vegetation structures, such as shrubs and vegetated tree grates or phenological characteristics, such as foliage on the street can be identified (refer to Figure 3, basement, first, and second floor, simulation, respectively as well as Figure 4, first and second floor, masked simulation, respectively). The simulation of tree crowns demonstrates a high degree of accuracy; however, disparities can be identified due to the exclusion of vegetation growth periods from the simulation process (refer to subsection 2.1 and Figure 3, basement floor, simulation, tree in the middle and first floor, simulation, right tree).

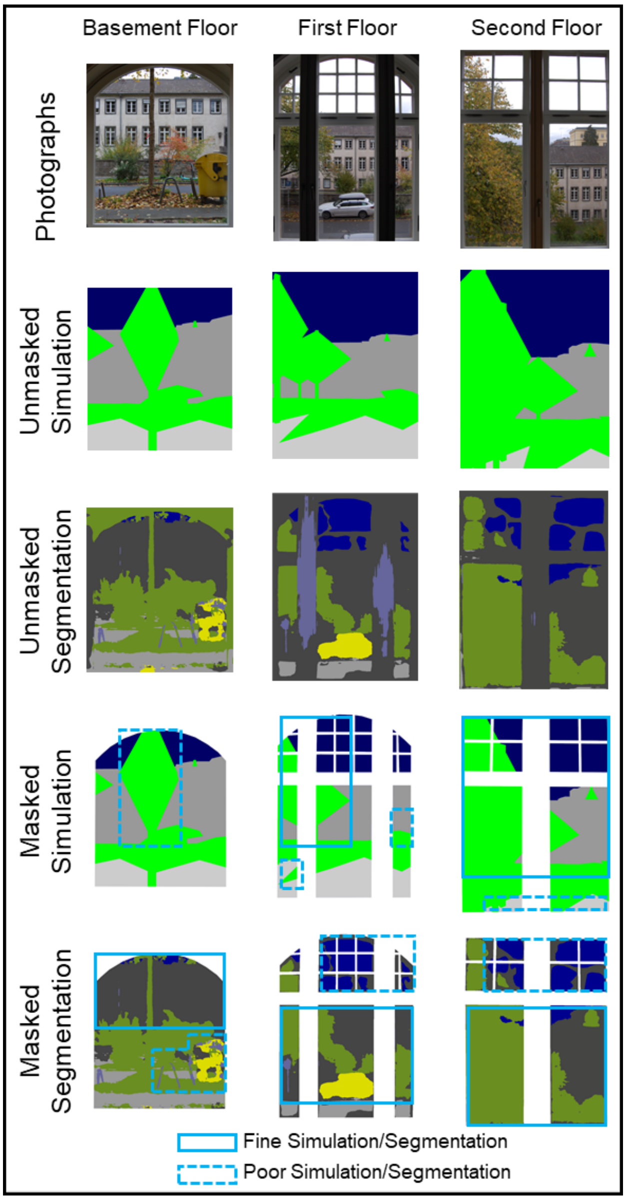

Figure 4. Exemplary comparison of masked window view simulation to masked photorealistic semantic segmentation of Deeplab V3+ pretrained on Cityscapes at a window distance of 2.0 m.

3.2 Intersection quality of window view simulation

As the visual comparison suggests, the metric values show a wide variety of overlap results depending on the height and position of the windows. (refer to Table 4 and 5).

Floors	OAcc	mAcc	mIoU	IoU_{veg}	IoU_{sky}	IoU_{buil}
Basement floor	0.64	0.69	0.45	0.47	0.52	0.36
First floor	0.67	0.74	0.47	0.47	0.52	0.42
Second floor	0.77	0.82	0.62	0.58	0.70	0.59

Table 4. Overall validation metrics and per-class IoU for window views at a window distance of 0.1 m.

Floors	OAcc	mAcc	mIoU	IoU_{veg}	IoU_{sky}	IoU_{buil}
Basement floor	0.53	0.67	0.28	0.32	0.18	0.32
First floor	0.57	0.63	0.53	0.50	0.87	0.23
Second floor	0.78	0.78	0.64	0.63	0.83	0.46

Table 5. Overall validation metrics and per-class IoU for window views at a window distance of 2.0 m.

The IoU for vegetation is 0.52. Sky reaches an IoU of 0.64 and the IoU of buildings is 0.43. The mIoU value is 0.53 and the mAcc reaches 0.74, indicating good class accuracy. The OAcc reaches a value of 0.68

3.3 Visibility values of window view simulation

The deviations of the visible labels based on the window view simulation and the semantic segmentation are low on average. However, they vary depending on the height and position of the windows (refer to Table 6 and 7).

Floors	$WVI\text{-}Diff_{veg}$	$WVI\text{-}Diff_{sky}$	$WVI\text{-}Diff_{buil}$
Basement floor	-0.05	0.05	0.08
First floor	-0.07	0.09	0.05
Second floor	-0.18	0.07	0.08

Table 6. Per-class visibility validation for window views at a window distance of 0.1 m.

Floors	$WVI\text{-}Diff_{veg}$	$WVI\text{-}Diff_{sky}$	$WVI\text{-}Diff_{buil}$
Basement floor	0.06	0.08	-0.01
First floor	0.01	0.01	-0.08
Second floor	0.02	-0.01	-0.01

Table 7. Per-class visibility validation for window views at a window distance of 2.0 m.

On average, there is an underestimation of vegetation (-6%), while the sky (+5%) and the building (+3%) are overestimated.

4. Discussion

4.1 Interpretation of validation outcomes

In their study, Li et al. (2024) compared the performance of a CIM window view (CIM-WV) with Cityscapes dataset based on a 7-class semantic segmentation using DeepLab V3+ (backbone

= Xception, OS = 16, trained on ImageNet) (Li et al., 2024). In this study, the model trained on CIM-WV achieved an OAcc of 97.60%, an mAcc of 91.48%, and an mIoU of 76.78%. The per-class IoU values were also high, with an IoU of 85.23% for vegetation, 98.67% for sky, and 97.49% for buildings. In contrast, the model trained on Cityscapes achieved comparatively low overall metrics with OAcc of 59.18%, mAcc of 50.98%, and mIoU of 34.14%. The IoU of vegetation reached 40.90% (compared to 52% in our validation) and the IoU of sky was 94.25% (64%). Buildings reached an IoU of 53.43% (43%). Factors that contributed to the observed discrepancies included the use of different viewpoints of the datasets, namely the window and street perspectives, which led to inaccuracies in the segmentation process. (Li et al., 2024)

A recent study by Peng et al. (2025) examined 11 classes in approximately 10,000 window views of an online platform that offers virtual 3D indoor apartment tours. SegNeXt was used, which achieves better results than DeepLab V3+ when trained on Cityscapes datasets. By including data augmentation, the model achieved an mAcc of 86.33% and an mIoU of 78.71%. Results based on raw data achieved an mAcc of 68.72% and an mIoU of 59.17%. (Peng et al., 2025)

Due to the limited number of studies that have analyzed and evaluated automatic window view techniques using semantic segmentation a comprehensive comparison with our results is not possible. It should also be noted that our results are based on a pretrained segmentation model. Therefore, a clear comparison with existing methods is limited. However, the validation process yielded encouraging results, indicating the potential of the simulation engine to generate window views for various window sizes and distances to the window.

4.2 Limitations

The present validation serves to verify and demonstrate the feasibility of a window view simulation engine combining open source approaches with open geospatial data. Despite promising results, it is evident that the spatial and temporal resolution of the open datasets leads to a limited level of detail in the simulated window views. This results in a lack of detail in the built and unbuilt urban morphology (walls, street signs, vehicles) and limited representation of phenological features, small 3D green structures, and vegetation growth rates (Bolte et al., 2019). To address these limitations, it is recommended to use vegetation models that are robust for modeling urban vegetation structures (Münzinger et al., 2022). In addition, more detailed 3D modeling of built-up urban morphology and street canyons based on airborne, terrestrial or mobile laser scanning surveys should be considered. However, these are comparatively time-consuming, labor-intensive, and expensive for large-scale, city-wide analyses. (Yu et al., 2024)

Due to the limited accessibility of indoor areas for documenting window views, this validation was conducted on a limited number of 40 windows. Consequently, the findings are not universally applicable to buildings with a different purposes or urban density values. It is recommended that a large-scale study be performed using crowdsourcing methods, including the use of mobile phone applications or questionnaires, to obtain a significant number of classified window views for ground truth (Bolte et al., 2023; 2025).

5. Conclusion

A window view simulation engine around the GWVI that incorporates open source techniques and open geospatial data was validated, resulting in an mIoU of 0.53 and an OAcc of 0.68, indicating a promising degree of accuracy. The simulation effectively captures the urban morphology, yet it overestimates the sky and buildings while underestimating vegetation. The simulation engine demonstrates considerable potential for extensive urban green space evaluations. Future investigations should employ crowdsourcing techniques and higher-resolution datasets to enhance validation.

Acknowledgements

The authors gratefully acknowledge the open datasets from the Regional Government of Cologne, Geobasis NRW, and Copernicus Data Space Ecosystem. The authors would also like to thank the reviewers for their helpful comments.

References

Amerio, A., Brambilla, A., Morganti, A., Aguglia, A., Bianchi, D., Santi, F., Costantini, L. et al., 2020: COVID-19 Lockdown: Housing Built Environment's Effects on Mental Health. *International journal of environmental research and public health*, 17(16): 5973–82. doi:10.3390/ijerph17165973

Basu, R., Colaninno, N., Alhassan, A., Sevtsuk. A., 2024: Hot and Bothered: Exploring the Effect of Heat on Pedestrian Route Choice Behavior and Accessibility. *Cities,* 155 (December 2024): 105435. doi.org/10.1016/j.cities.2024.105435

Bolte, A.-M., Kötter, T., Schuppe, S., 2019. Can you see green or blue? On the necessity of visibility analysis of urban open spaces using Remote Sensing techniques and Geographic Information Systems. 2019 Joint Urban Remote Sensing Event (JURSE) May, 22-24, 2019, Vannes, France. 1-4. doi.org/10.1109/JURSE.2019.8808936

Bolte, A.-M., Moghadas, M., Kötter, T., 2023. Harnessing Crowdsourcing Data for Comprehensive Green Window View Analysis. LET IT GROW, LET US PLAN, LET IT GROW. Nature-based Solutions for Sustainable Resilient Smart Green and Blue Cities. Proceedings of REAL CORP 2023, 28th International Conference on Urban Development, Regional Planning and Information Society. 351-360. doi.org/10.48494/REALCORP2023.8042

Bolte, A.-M., Niedermann, B., Kistemann, T., Haunert, J.-H., Dehbi, Y., Kötter, T., 2024a: The green window view index: automated multi-source visibility analysis for a multi-scale assessment of green window views. *Landscape Ecology*, 39/71, 1-26. doi.org/10.1007/s10980-024-01871-7.

Bolte, A.-M., Olefs, M., Kötter, T., 2024b. Wie viel Grün sieht die Stadt? Automatisierte mehrskalige Sichtbarkeitsanalyse von städtischen Grünflächen einer nordrhein-westfälischen Großstadt. (How much green does the city see? Automated multi-scale visibility analysis of urban green spaces in a major city in North Rhine-Westphalia). Flächennutzungsmonitoring XVI – Flächenpolitik – Flächenanalysen – Methoden und Werkzeuge IÖR Schriften Band 82, 2024, 237-246. doi.org/10.5281/zenodo.14699410

Bolte, A.-M., Kistemann, T., Dehbi, Y., Kötter, T., 2025: (Un)just distribution of visible green spaces? A socio-economic window view analysis on residential buildings: The City of Cologne as case study. *Journal of Geovisualization and Spatial Analysis* xx/xx, xx-xx. doi:10.1007/s41651-025-00214-7

Bundesamt für Kartographie und Geodäsie (BKG), 2024. Vertrag über die kontinuierliche Übermittlung amtlicher digitaler Geobasisdaten der Länder zur Verwendung im Bundesbereich (V GeoBund 2024). (Contract on the continuous transmission of official digital geodata of the federal states for use in the federal sector).

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham. doi.org/10.1007/978-3-030-01234-2_49

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi.org/10.1109/CVPR.2016.350

Fang, G, 2020. DeepLabv3Plus-Pytorch. Pretrained DeepLabv3, DeepLabv3+ for Pascal VOC & Cityscapes. (2020)

Haaland, C., Konijnendijk van den Bosch, C., 2015: Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban Forestry & Urban Greening*, 14(4), 760–771. doi.org/10.1016/j.ufug.2015.07.009

Li, M., Xue, F.m Wu, Y., Yeh, A. G. O., 2022: A room with a view: Automatic assessment of window views for high-rise high-density areas using City Information Models and deep transfer learning. *Landscape and Urban Planning*, 226 (2022), 1-13. doi.org/10.1016/j.landurbplan.2022.104505

Li, M., Yeh, A. G. O., Xue, F., 2024: CIM-WV: A 2D semantic segmentation dataset of rich window view contents in high-rise, high-density Hong Konk based on photorealistic city information models. *Urban Informatics*, 3/12, 1-19. doi.org/10.1007/s44212-024-00039-7

Meng, X., Wang, M., 2024: Exploring the health impacts if window views: a literature review. *Journal of Asian Architecture and Building Engineering*, 1-24. doi.org/10.1080/13467581.2024.2412120

Münzinger, M., Prechtel, N., Behnisch, M., 2022: Mapping the urban forest in detail: From LiDAR point clouds to 3D tree models. *Urban Forestry & Urban Greening,* 74 (August 2022), 1-13. doi.org/10.1016/j.ufug.2022.127637

Peng, C., Xiang, Y., Huang, W., Feng, Y., Tang, Y., Biljecki, F., Zhou, Z., 2025: Measuring the value of window views using real estate big data and computer vision: A case study in Wuhan, China. *Cities*, 156 (2025), 1-17. doi.org/10.1016/j.cities.2024.105536

Pijpers, R., van Melik, R., 2020. Looking, Observing, Watching: Meaningful Contact at a Distance. In Age-Inclusive Public Spaces, Altenburg: Hatje Cantz Verlag GmbH, 237.

Swietek, R., Zumwald, M., 2023: Visual Capital: Evaluating building-level visual landscape quality at scale. *Landscape and Urban Planning*, 240 (2023), 1-14. doi.org/10.1016/j.landurbplan.2023.104880

Yu, B., Wang, Y., Chen, Q., Chen, X., Zhang, Y., Luan, K., Ren, X., 2024: A review of road 3D modeling based on light detection and ranging point clouds. *Journal of Road Engineering*, 4/4, 386-398. doi.org/10.1016/j.jreng.2024.04.009