
Facade Segmentation for Solar Photovoltaic Suitability

Ayca Duran

ETH Zurich
Architecture and Building Systems
Zurich, Switzerland
duran@arch.ethz.ch

Christoph Waibel

VITO
CRIB
Mol, Belgium
christoph.waibel@vito.be

Bernd Bickel

ETH Zurich
Computational Design Lab
Zurich, Switzerland
bickelb@ethz.ch

Iro Armeni

Stanford University
Gradient Spaces Lab
Stanford, CA, US
iarmeni@stanford.edu

Arno Schlueter

ETH Zurich
Architecture and Building Systems
Zurich, Switzerland
schlueter@arch.ethz.ch

Abstract

Building integrated photovoltaic (BIPV) facades represent a promising pathway towards urban decarbonization, especially where roof areas are insufficient and ground-mounted arrays are infeasible. Although machine learning-based approaches to support photovoltaic (PV) planning on rooftops are well researched, automated approaches for facades still remain scarce and oversimplified. This paper therefore presents a pipeline that integrates detailed information on the architectural composition of the facade to automatically identify suitable surfaces for PV application and estimate the solar energy potential. The pipeline fine-tunes SegFormer-B5 on the CMP Facades dataset and converts semantic predictions into facade-level PV suitability masks and PV panel layouts considering module sizes and clearances. Applied to a dataset of 373 facades with known dimensions from ten cities, the results show that installable BIPV potential is significantly lower than theoretical potential, thus providing valuable insights for reliable urban energy planning. With the growing availability of facade imagery, the proposed pipeline can be scaled to support BIPV planning in cities worldwide.

1 Introduction

The energy use of buildings during their operation accounts for nearly 28% of global carbon emissions [1]. Building-integrated photovoltaics (BIPV) offer a promising pathway to reduce these emissions, with cost reductions and technological advances driving a wider adoption. Although rooftop photovoltaic (PV) design has been extensively researched, automated approaches to facade photovoltaic energy remain underexplored despite its relevance to urban renewable energy [2, 3]. Especially in high-rise and mixed-use developments, where roof space is limited and ground-mounted systems are constrained by land availability or regulations, BIPV can be a viable design choice. However, barriers to facade BIPV integration include current decision support workflows that are slow and tedious, expert-dependent, and difficult to scale, thus delaying a wider adoption despite the benefits [2].

A range of Geographic Information System (GIS)-enabled tools exist for PV potential assessment, which typically rely on rooftop geometries, satellite imagery, and irradiance databases, to identify installation areas and estimate energy yields. Some extend to facade applications, such as Sonnenfassade.ch [4] and PVWatts [5], but these often treat facades as uniform surfaces and overlook their complex architectural qualities. Although archetype-based approaches exist [6], a thorough consideration of facade-level details and obstructions remains an understudied area. Recent work

has applied conditional adversarial networks [7] to segment facade imagery using the CMP Facades dataset [8] and estimated the theoretical PV potential assuming that the raw wall share of a facade can be fully covered with BIPV panels [9]. Although this approach automatically identifies PV-eligible wall surfaces, it does not progress to practical system design, such as feasible panel layouts.

In this study, we introduce a segmentation-and-layout workflow that bridges the gap between identifying eligible surfaces and a practical system design. This automated workflow includes (1) segmenting facade components, (2) constructing PV-suitability masks by excluding openings and protrusions with clearance buffers, and (3) generating practical panel layouts based on module size. This produces estimates in physical units (area in m^2 , peak power in kWp, and annual energy yield in kWh) and layout visualizations suitable for integration into solar cadastres or retrofit planning.

2 Data and Method

Datasets. We use two datasets for model training and application. The CMP Facades dataset [8] provides annotations per pixel for 13 classes: *background*, *facade*, *window*, *door*, *cornice*, *sill*, *balcony*, *blind*, *molding*, *deco*, *pillar*, *shop* and *unknown*. The data set consists of 606 annotated images. The *unknown* class is ignored in both loss computation and metric evaluation. To demonstrate the conversion from pixel fractions to energy yield calculations, 373 facades are randomly sampled from the Large Scale Architectural Asset (LSAA) dataset [10]. In the application phase, ortho-rectified images and location data provided in the LSAA dataset are utilized. The dimensions and orientation of the facade are manually obtained from Google Earth, a platform easily accessible to users.

Model and implementation details. We adapt the classifier head of SegFormer-B5 model [11], initialized from the ADE20K[12]-finetuned checkpoint, to the 13 classes of CMP Facade dataset. The images are resized to 640×640 pixels. Data augmentation includes horizontal flips, small shifts, scaling and rotation, brightness/contrast adjustment, HSV jitter, and light Gaussian blur (Appendix A.2). The model is trained using the AdamW optimizer (learning rate 2×10^{-4} , weight decay 10^{-2}) with a cosine learning rate schedule and 5% warm-up. The batch size is 4, constrained by GPU memory. The loss combines inverse-frequency class-weighted cross-entropy and $0.5 \times$ soft Dice loss over non-*unknown* classes. Early stopping (patience 10) is applied within 80 epochs. At inference, logits are bilinearly upsampled and argmax operation is applied per pixel. The hyperparameters are tuned following a small grid-search (Appendix A.3).

PV mask construction. PV-suitable masks are derived from semantic predictions by (i) selecting pixels classified as *facade*, (ii) removing *window*, *door*, and *shop* pixels with dilated buffers to account for frames and installation clearances, (iii) excluding *balcony*, *cornice*, *sill*, *molding*, *deco*, *pillar*, and *blind* with class-specific buffers and area thresholds, and by (iv) removing small connected components below a minimum area threshold equivalent to one panel size with buffers.

Let P denote PV-suitable pixels and F all facade pixels. The PV fraction per image is then:

$$f_{\text{PV}} = \frac{|P|}{|F|}. \quad (1)$$

With known facade area A_{facade} (m^2), the PV-suitable area is:

$$A_{\text{PV}} = f_{\text{PV}} \times A_{\text{facade}} \quad (\text{m}^2). \quad (2)$$

Panel layout generation and energy yield. To translate PV-suitable areas into installable designs, we fit standard PV module grids inside the predicted mask. The largest axis-aligned rectangle fully contained within the mask is then located using a histogram-based maximal-rectangle search. Within this rectangle, we test both portrait and landscape orientations of common *large* (L) and *small* (S) module sizes with a fixed inter-panel gap (Appendix A.4). The number of modules along each axis is computed by the integer division of the rectangle size by the module-plus-gap footprint. The orientation that maximizes the count while minimizing leftover space is selected.

Given a rating for standard test conditions (STC) per module $P_{\text{module,STC}}$ and the fit module count N_{modules} , the installed peak DC capacity is

$$P_{\text{DC,STC}} = P_{\text{module,STC}} \times N_{\text{modules}} \quad (\text{kWp}), \quad (3)$$

Method	mIoU	Pixel Accuracy	Macro Precision	Macro Recall	Macro F1
Ours (SegFormer-B5)	0.52	0.76	0.68	0.68	0.68
Auto-Context (ST3) [14]	0.36	0.66	—	0.49	—
Auto-Context (PW3) [14]	0.38	0.68	—	0.49	—
Baseline – Uniform Random	0.03	0.08	0.08	0.06	0.05
Baseline – Majority Class	0.03	0.24	0.02	0.08	0.03

Table 1: Segmentation performance on the CMP Facades. (–: unavailable metric in the cited study.)

Method	Bg	Fac	Win	Door	Corn	Sill	Balc	Blind	Mold	Deco	Pill	Shop
Ours	0.74	0.62	0.69	0.46	0.51	0.48	0.50	0.56	0.37	0.52	0.33	0.43

Table 2: Per-class IoU on CMP Facades dataset.

We also calculate the location- and orientation-specific annual energy yield ($E_{PV,ann}$) using PVGIS24 [13]. The details of the calculation can be found in the Appendix A.5. The code is available on GitHub¹.

3 Experiments and Results

Setup and evaluation metrics. We use the CMP Facades dataset for training, validation, and testing [8]. Segmentation performance for 12 classes is reported using dataset-level mean Intersection-over-Union (mIoU) and mean over per-image mIoU values, as well as pixel accuracy, macro precision, recall, and F1-score over all classes. The formal definitions of these metrics are provided in Appendix A.1. Experiments ran on a Windows workstation (Ryzen 9 7950X; RTX 4090 24 GB; driver 560.94/CUDA 12.6).

Segmentation performance. Our SegFormer-B5 outperforms classical Auto-Context baselines on CMP. Dataset-level mIoU is 0.52 (per-image mean mIoU 0.44), with macro precision/recall/F1 of 0.68/0.68/0.68. To contextualize difficulty, we include two baselines: Uniform Random (each pixel sampled uniformly from the 12 evaluated classes) and Majority Class (all pixels predicted as background), evaluated with the same metrics. The IoU score per class is highest for large, distinct classes such as *background* and *window*, and lowest for small, diverse elements such as *pillar* (Table 2). We attribute this gap to dataset class imbalance and resolution effects from the stride-4 decoder and bilinear upsampling, although we achieved performance improvements in smaller classes with data augmentation efforts.

Class-share on the facade surface. PV suitability of a facade is commonly reported as a fraction of facade area [9]. Therefore, we additionally present a domain-specific metric that compares the predicted and true shares of three class groups relative to facade surface: *wall*, glazing (*window+shop*), and others. The predicted pixel share for major facade groups closely matches the ground truth (GT): *wall* 47.42% (GT: 49.39%), *windows+shop* 16.48% (GT: 16.08%), other classes 36.10% (GT: 34.53%). The wall share error is -1.97 percentage points (pp). Across individual images, the mean absolute error is 7.29 pp and the root mean square error is 12.17 pp.

Application to LSAA. The trained pipeline is applied to 373 facades from the LSAA dataset with known facade areas. For each facade, PV fraction f_{PV} and PV-suitable area A_{PV} as well as the largest panel layout within the PV-suitable area are calculated. The method yields an average PV-suitable fraction of around 38–39% of the predicted facade area after applying clearance buffers and component cleaning depending on the module size (Table 3). Given the known facade and panel dimensions, the fitted layout was converted to installed peak capacity (kWp). Qualitative examples of the resulting layouts is presented in Figure 1 and discussed further in Appendix A.6, which illustrates common failure modes related to occlusions, segmentation accuracy, and facade complexity.

Theoretical vs practical panel placement. Figure 1 shows the outputs of each step in the proposed workflow for eight example facades from different cities. Applying the panel layout generation procedure reveals cases where large PV-suitable areas in the binary mask do not translate into practical systems. In total, 14 facades with $f_{PV} > 0.30$ yield a panel count of zero due to high

¹<https://github.com/ycdrn/segment4pvlayout>

Wall surface (%)	Theoretically suitable (%)	Practically suitable (%)
54.01	38.62–39.48	4.35–4.75

Table 3: Average shares of PV-suitable surfaces (*large–small* panels) for the LSAA subset.

ornamentation and narrow or irregular mask regions that cannot fit a single full-size S module with clearances. The fraction of facade area that can be practically installed with solar panels (L – S size) range between 4 – 5%, compared to a mean theoretical PV-suitable share of around 39% across the LSAA subset (Table 3). The results confirm that relying on theoretical availability can overestimate installable capacity, particularly for highly articulated facades, as observed in the application dataset representing the old building stock in different cities.

City	Vienna	Brussels	Copenhagen	Paris	Rome	Amsterdam	London	Chicago
Segmentation								
Theoretical P.								
Practical P.								
γ	–38°	146°	–80°	–25°	–66°	–91°	13°	–1°
P_{DC-STC}	9.45 kWp	3.52 kWp	11.48 kWp	4.5 kWp	4.05 kWp	2.7 kWp	2.48 kWp	6.75 kWp
$E_{PV,ann}$	8,182.66 kWh	1014.66 kWh	7122.42 kWh	3945.15 kWh	3740.82 kWh	1509.49 kWh	2300.41 kWh	7235.39 kWh

Figure 1: Example applications (with panel size L) from the LSAA subset ($N = 373$).

4 Conclusion

This study introduces a workflow that generates PV-suitability masks from facade images and translates them into practical panel layouts and energy estimates. By combining semantic segmentation and computational design methods, the workflow moves beyond theoretical potential estimation to produce layouts that can potentially be installed in practice. Evaluation on the CMP Facades dataset shows competitive performance across major facade classes, and application to a subset of the LSAA dataset demonstrates scalability across cities. By automatically estimating A_{PV} for large building stocks, the approach helps prioritize retrofits toward facades with large, uninterrupted wall fractions. When converted to kWp and paired with irradiance profiles, the results can support distribution-network planning by mapping potential generation in spatial and temporal dimensions.

The results for the PV layouts are based on a common L and S BIdPV panel sizes for facades. Varied or custom-sized panels could increase practical potential but would likely incur higher upfront costs. The generated PV layouts represent the maximum semantically suitable surfaces on facades within the given geometric constraints, excluding potential aesthetic considerations. Our method generates PV panel layouts using images and PVGIS integration to predict maximum energy yields by considering site and orientation specific energy yields. However, if 3D models are available, the accuracy of energy yield estimations can be improved through radiation simulations, particularly for surfaces that are substantially shaded. The results support early stage decision making rather than permitting decisions. Future work will focus on evaluating the generated PV layouts, improving the accuracy of segmentation performance, and generating design alternatives beyond the maximum panel coverage to help decision making in facade retrofits with BIPV.

References

- [1] United Nations Environment Programme. *2023 Global Status Report for Buildings and Construction: Beyond foundations - Mainstreaming sustainable solutions to cut emissions from the buildings sector*. United Nations Environment Programme, March 2024. ISBN 978-92-807-4131-5. doi: 10.59117/20.500.11822/45095. URL <https://wedocs.unep.org/20.500.11822/45095>.
- [2] Antonin Faes, Alessandro Virtuani, Hugo Quest, Laura Maturi, Alessandra Scognamiglio, Francesco Frontini, Arno Schlueter, Nuria Martin-Chivelet, Angèle Reinders, and Christophe Ballif. Building-integrated photovoltaics. *Nature Reviews Clean Technology*, 1(5):333–350, May 2025. ISSN 3005-0685. doi: 10.1038/s44359-025-00059-9. URL <https://www.nature.com/articles/s44359-025-00059-9>.
- [3] Ayca Duran, Christoph Waibel, Valeria Piccioni, Bernd Bickel, and Arno Schlueter. A review on artificial intelligence applications for facades. *Building and Environment*, 269:112310, February 2025. ISSN 03601323. doi: 10.1016/j.buildenv.2024.112310. URL <https://linkinghub.elsevier.com/retrieve/pii/S0360132324011521>.
- [4] Markus Portmann, David Galvagno-Erny, Priska Lorenz, David Schacher, and Rolf Heinrich. Sonnen-dach.ch und Sonnenfassade.ch: Berechnung von Potenzialen in Gemeinden. Technical report, Bern, 2019.
- [5] A. Dobos. PVWatts Version 5 Manual. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO (United States), September 2014. URL <http://www.osti.gov/servlets/purl/1158421/>.
- [6] Erika Saretta, Pierluigi Bonomo, and Francesco Frontini. A calculation method for the BIPV potential of Swiss façades at LOD2.5 in urban areas: A case from Ticino region. *Solar Energy*, 195:150–165, January 2020. ISSN 0038092X. doi: 10.1016/j.solener.2019.11.062. URL <https://linkinghub.elsevier.com/retrieve/pii/S0038092X19311624>.
- [7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks, 2016. URL <https://arxiv.org/abs/1611.07004>. Version Number: 3.
- [8] Radim Tyleček and Radim Šára. Spatial Pattern Templates for Recognition of Objects with Regular Structure. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *Pattern Recognition*, volume 8142, pages 364–374. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40601-0 978-3-642-40602-7. doi: 10.1007/978-3-642-40602-7_39. URL http://link.springer.com/10.1007/978-3-642-40602-7_39. Series Title: Lecture Notes in Computer Science.
- [9] Ayca Duran, Christoph Waibel, and Arno Schlueter. Estimating surface utilization factors for BIPV applications using pix2pix on street captured façade images. *Journal of Physics: Conference Series*, 2600(4):042005, November 2023. doi: 10.1088/1742-6596/2600/4/042005. URL <https://dx.doi.org/10.1088/1742-6596/2600/4/042005>. Publisher: IOP Publishing.
- [10] Peihao Zhu, Wamiq Reyaz Para, Anna Fruhstuck, John Femiani, and Peter Wonka. Large-Scale Architectural Asset Extraction from Panoramic Imagery. *IEEE Transactions on Visualization and Computer Graphics*, 28(2):1301–1316, February 2022. ISSN 1077-2626, 1941-0506, 2160-9306. doi: 10.1109/TVCG.2020.3010694. URL <https://ieeexplore.ieee.org/document/9145640/>.
- [11] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, 2021. URL <https://arxiv.org/abs/2105.15203>. Version Number: 3.
- [12] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene Parsing through ADE20K Dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5122–5130, 2017. doi: 10.1109/CVPR.2017.544.
- [13] Thomas Huld, Richard Müller, and Attilio Gambardella. A new solar radiation database for estimating PV performance in Europe and Africa. *Solar Energy*, 86(6):1803–1815, June 2012. ISSN 0038092X. doi: 10.1016/j.solener.2012.03.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S0038092X12001119>.
- [14] Raghudeep Gadde, Varun Jampani, Renaud Marlet, and Peter V. Gehler. Efficient 2D and 3D Facade Segmentation using Auto-Context, 2016. URL <https://arxiv.org/abs/1606.06437>. Version Number: 1.

A Appendix

A.1 Metric Definitions

This appendix defines the evaluation metrics used in Section 3, which quantify different aspects of segmentation accuracy, class balance, and pixel-share estimation. Let TP_c , FP_c , and FN_c denote the true positive, false positive, and false negative pixel counts for class c , and let C be the set of evaluated classes (all except *unknown*).

Intersection-over-Union (IoU), measures the overlap between predicted and ground truth regions for class c , relative to their union:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c}. \quad (4)$$

Mean Intersection-over-Union (mIoU), summarizes overall segmentation performance across all classes:

$$mIoU = \frac{1}{|C|} \sum_{c \in C} IoU_c. \quad (5)$$

Two variants are reported: (1) dataset-level mIoU, computed from aggregated counts across all test images; and (2) mean over per-image mIoU values.

Macro Precision and Recall, indicate, respectively, the proportion of predicted pixels that are correct and the proportion of ground truth pixels that are recovered, averaged equally over classes. Macro recall corresponds to ‘‘Average class accuracy’’ reported in [14].:

$$\text{Precision} = \frac{1}{|C|} \sum_{c \in C} \frac{TP_c}{TP_c + FP_c}, \quad (6)$$

$$\text{Recall} = \frac{1}{|C|} \sum_{c \in C} \frac{TP_c}{TP_c + FN_c}. \quad (7)$$

Macro F1 score is the harmonic mean of macro precision and recall, giving a single measure of class-balanced accuracy:

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (8)$$

Pixel Accuracy (Overall) is the proportion of correctly classified pixels over all non-ignored pixels:

$$\text{Acc} = \frac{\sum_{c \in \tilde{C}} TP_c}{\sum_{c \in \tilde{C}} (TP_c + FN_c)}$$

where $\tilde{C}$ denotes all non-ignored labels (including class 0). This corresponds to ‘‘Overall’’ in [14].

Class-share consistency evaluates how well the predicted pixel share of each class matches the ground truth. For image i , let $s_{c,i}$ be the predicted share of pixels for class c and $s_{c,i}^{\text{GT}}$ the ground-truth share. The signed error for class c is

$$\Delta s_{c,i} = s_{c,i} - s_{c,i}^{\text{GT}} \quad (\text{pp}),$$

where ‘‘pp’’ denotes percentage points. The mean absolute error (MAE) and root-mean-square error (RMSE) across N images are:

$$\text{MAE}_c = \frac{1}{N} \sum_{i=1}^N |\Delta s_{c,i}|, \quad (9)$$

$$\text{RMSE}_c = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta s_{c,i})^2}. \quad (10)$$

A.2 Data Augmentation Details

All enhancements are applied on the fly during training at a resolution of 640×640 . Transformations include horizontal flips with probability $p = 0.5$; random shifts of up to $\pm 10\%$, scaling between 0.9 and 1.1, and rotations of up to $\pm 5^\circ$; brightness and contrast adjustments with factors in $[0.9, 1.1]$; HSV jitter with hue shifts of $\pm 10^\circ$ and saturation and value scaling in $[0.9, 1.1]$; and Gaussian blur with a kernel of 3×3 and $\sigma \in [0.0, 1.0]$.

A.3 Hyperparameter Tuning

To set optimizer and training hyperparameters, we carried out a small grid search on the CMP validation split. For each configuration, we trained the SegFormer model initialized from ADE20K for up to 80 epochs with a cosine

learning-rate schedule and linear warmup (warmup ratio 0.05), using early stopping when the validation mIoU did not improve for 10 consecutive epochs. The search space comprised learning rates $\{5 \times 10^{-5}, 10^{-4}, 2 \times 10^{-4}\}$, batch sizes $\{2, 4\}$, and gradient accumulation steps $\{1, 2\}$, with weight decay fixed at 10^{-2} . The final model uses the configuration that achieved the highest validation mIoU under the sweep: learning rate 2×10^{-4} , weight decay 10^{-2} , warmup ratio 0.05, batch size 4, and accumulation steps 1.

A.4 Panel Placement Parameters

The placement of modules is performed for *large* (935×1300 mm) and *small* (720×875 mm) facade modules² with a fixed gap between panels of 0.02 m. Both portrait and landscape orientations are tested.

The binary PV mask has image dimensions $(W_{\text{px}}, H_{\text{px}})$ and known facade size $(W_{\text{m}}, H_{\text{m}})$. The scale factors are

$$s_x = \frac{W_{\text{px}}}{W_{\text{m}}}, \quad s_y = \frac{H_{\text{px}}}{H_{\text{m}}},$$

allowing panel dimensions and gaps to be expressed in pixels.

To locate the largest axis-aligned rectangle within the mask, we build row-wise histograms of consecutive “PV” pixels per column and compute the maximal rectangle per histogram using a standard stack-based algorithm. This ensures the rectangle is fully contained in the PV region.

Within the rectangle, the number of modules per axis is computed as

$$n_x = \left\lfloor \frac{R_w + g}{p_w + g} \right\rfloor, \quad n_y = \left\lfloor \frac{R_h + g}{p_h + g} \right\rfloor,$$

where (R_w, R_h) are the rectangle dimensions in pixels, (p_w, p_h) the panel size in pixels, and g the inter-panel gap in pixels. Panels are accepted only if fully contained; partial panels are excluded. The resulting grid is centered in the rectangle.

A.5 Energy Yield Estimation

The annual energy yield is computed from the installed peak DC capacity depending on the module type (*large* = 225 Wp , *small* = 110 Wp)² and the PVGIS-returned annual specific yield:

$$E_{\text{PV,ann}} = P_{\text{DC,STC}} \times Y_{\text{spec}} \quad (\text{kWh/yr}), \quad (11)$$

where $P_{\text{DC,STC}}$ is in Wp (from Equation 2) and Y_{spec} is the location- and orientation-specific annual specific yield $[\text{kWh/kWp/yr}]$ returned by PVGIS given the inputs (tilt 90° , facade azimuth, integrated mounting, and default loss parameters: cable 1%, inverter 2%, PV 0.5%).

$$Y_{\text{spec}} = \text{PR}_{\text{ann}} \times \frac{H_{\text{POA,ann}}}{G_{\text{STC}}}, \quad G_{\text{STC}} = 1 \text{ kW/m}^2, \quad (12)$$

with $H_{\text{POA,ann}}$ the annual plane-of-array irradiation $[\text{kWh/m}^2/\text{yr}]$ and PR_{ann} the annual performance ratio (%) accounting for temperature, spectral and angle-of-incidence effects, and system losses as modeled by PVGIS. Substituting Equation 12 into Equation 11 yields:

$$E_{\text{PV,ann}} = P_{\text{DC,STC}} \text{PR}_{\text{ann}} \frac{H_{\text{POA,ann}}}{G_{\text{STC}}}, \quad (13)$$

A.6 Common Failure Modes in Layout Generation

A qualitative assessment shows three recurring failure modes (Fig.2a–c). Foreground occlusions, such as trees, people, vehicles, trees, are not explicitly segmented and are thus treated as walls, resulting in unreliable PV areas (Fig.2a). This can be mitigated by an occlusion-removal step, such as detecting and masking foreground objects or inpainting, and by adding an explicit *occlusion/vegetation* class, especially for trees that may cast shade. Pitched roofs are another source of error due to limited representation in CMP, leading to poor segmentation on sloped geometries (Fig.2b). Adding pitched-roof examples and marking such these as non-suitable can improve robustness. Finally, highly articulated or repetitive facades, such as textured surfaces, and arrays of balconies, reduce segmentation quality and expose a limitation of the largest-rectangle heuristic, which overlooks multiple smaller feasible regions (Fig.2c). Allowing multi-region packing, in other words placing panels across several disjoint areas, relaxing the single-rectangle assumption, and enforcing minimum cluster sizes can yield layouts with larger energy generation capacity.

²<https://www.3s-solar.swiss/downloads-teraslate#systemeproduktes>

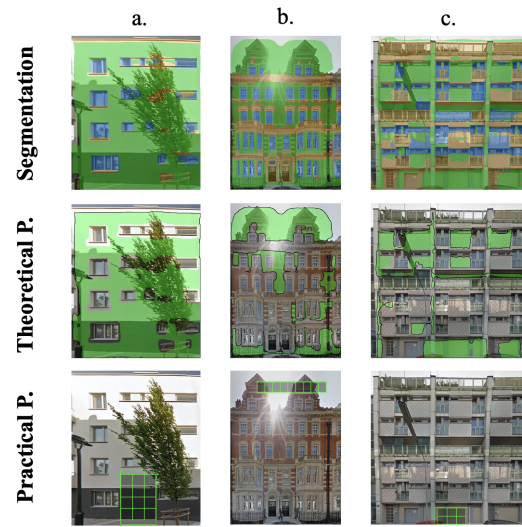


Figure 2: Failure modes due to occlusions (a), segmentation accuracy (b), complexity of facade (c).