Large-Scale Rooftop Solar Energy Estimation Using Deep Learning and Aerial Imaging

Stanley Egbe
SOE107@student.aru.ac.uk
Mahdi Maktab Dar Oghaz
mahdi.maktabdar@aru.ac.uk
Lakshmi Babu Saheer
lakshmi.babu-saheer@aru.ac.uk

Faculty of Science and Engineering
Anglia Ruskin University
Cambridge, UK

Abstract

Cities need credible, comparable maps of rooftop photovoltaic (PV) potential to support net-zero planning, asset targeting, and grid coordination. We present a data-light pipeline that converts aerial imagery into planning-grade indicators by combining attention-augmented U-Net segmentation with a principled pixel-to-ground scaling procedure and a simple, orientation-aware energy model. Trained on INRIA aerial tiles, our Full-Attention U-Net achieves a test IoU of 0.874, outperforming baseline and skip-attention variants. Converting masks to square metres using tile-wise metres-per-pixel, we validate areas against Google Earth measurements: mean error is 3.23% on georeferenced dataset crops and 11.02% on Google Static Maps tiles. The novelty of this work lies in the integration of three elements: (i) a Full-Attention U-Net that improves segmentation fidelity, (ii) an explicit pixel-to-ground scaling procedure that provides robust geospatial accuracy, and (iii) the translation of rooftop masks into planning-grade PV indicators such as installable capacity, indicative yield, and ward-level aggregates. Beyond accuracy, the framework is intentionally data-light and scalable, remaining transferable across cities without LiDAR or dense cadastral data. This makes the approach highly policy-relevant, offering a pragmatic bridge between computer vision and urban sustainability practice. While current assumptions are 2D (no explicit tilt, shading, or superstructure masking), the approach offers a scalable bridge from computer vision to urban sustainability practice. We outline extensions to incorporate 3D roof facets, shading, and uncertainty propagation, and to validate against operational PV generation.

1 Introduction

Rooftop solar is one of the fastest ways cities can decarbonise electricity, cut bills, and improve energy resilience without expanding land use. For planners and utilities, credible maps of rooftop photovoltaic (PV) potential enable evidence-based zoning, target setting for net-zero strategies, and grid reinforcement planning at street or feeder level [5, 11]. They also

support social policy prioritising installations on public buildings and social housing to address energy poverty and inform co-deployment with heat pumps and EV charging. Yet most municipalities still rely on coarse assumptions or costly building-by-building surveys, creating uncertainty in capacity forecasts, siting decisions, and business cases for local energy communities [6, 18].

Despite strong progress in computer vision for building and roof extraction, two gaps remain for real-world deployment. First, state-of-the-art (SOTA) rooftop segmentation models often optimise pixel metrics but stop short of producing planning-ready outputs i.e., roof-usable area, orientation-aware panel layouts, and kilowatt/kilowatt-hour estimates that local authorities and distribution system operators (DSOs) can act on [14, 20, 21]. Second, scalability and generalisation are under-discussed: many pipelines depend on dense cadastral data, lidar, or city-specific calibration, which limits portability across jurisdictions and imagery providers. As a result, PV potential studies are difficult to compare across wards or towns, and downstream artefacts (e.g., overestimation on complex roofs, inconsistent pixel-to-ground conversion) propagate into capacity and carbon-savings estimates used by policy makers [4, 14, 21].

We address these gaps with a modular, data-light pipeline that goes from aerial imagery to planning-grade PV indicators. Using attention-enhanced U-Net variants trained on high-resolution aerial data, we generate precise rooftop masks and convert them into geospatially consistent roof polygons [14, 15]. A post-processing stage estimates usable area after boundary setbacks and obstruction filtering, assigns feasible panel layouts by roof facet and aspect, and converts area to installable capacity (kWp) and expected annual yield (kWh) using location-specific irradiance factors. Crucially, we aggregate building-level results to planning geographies (e.g., LSOA/ward) to produce the products decision-makers need: (i) PV capacity/yield maps for scenario analysis; (ii) suitability scores to rank candidate public assets; (iii) feeder-level demand-offset layers for DSOs; and (iv) carbon-abatement estimates to track progress against net-zero plans. Validation against independent measurements demonstrates reliable area estimates and stable performance on unseen neighbourhoods, supporting transfer to new locales without extensive re-survey [7, 9].

The core novelty of this research lies not in a single component but in the integration of three innovations into one pipeline. First, we design a Full-Attention U-Net that enhances rooftop segmentation by sharpening boundaries and reducing background noise, outperforming both conventional U-Net and skip-attention variants. Second, we introduce an explicit, validated pixel-to-ground scaling procedure that converts segmented pixels into physically meaningful square-metre areas, overcoming a key limitation of many prior PV-mapping studies that assume fixed resolution. Third, rather than stopping at pixel-level metrics, we translate segmentation outputs into planning-grade PV indicators installable capacity, indicative yield, and ward-level aggregates providing outputs directly useful for energy planners and utilities.

The manuscript is organised as follows: Section 2 reviews related work on rooftop detection, pixel-to-ground scaling, and imagery-based PV assessment; Section 3 details our dataset, attention-augmented architectures, geospatial scaling, obstruction handling, and the energy-modelling stack; Section 4 reports segmentation accuracy, area fidelity, and planning-level indicators with case studies on siting, targeting, and grid coordination; and Section 5 concludes with implications for urban planning, key limitations, and future work, including high-resolution shading, uncertainty quantification, and socio-economic adoption scenarios.

2 Literature Review

2.1 Rooftop detection from overhead imagery

Deep learning has rapidly advanced rooftop and PV-array mapping in aerial/satellite imagery. City-scale pipelines use semantic segmentation to delineate rooftops and derive solar-ready area; for instance, Zhong et al. developed a deep-learning framework that reduced labeling effort and generalized across diverse districts while estimating Nanjing's rooftop capacity [23]. Size-aware models explicitly target small PV arrays, Wang et al.'s Rooftop PV Segmenter (RPS) refines multi-scale features to better separate small solar modules from roof background [19]. Recent transformer-based and self-/weakly-supervised approaches seek stronger cross-domain generalization and reduced annotation cost, while Mask2Former-style architectures have improved PV-module instance segmentation across heterogeneous resolutions [8]. Public datasets and active-learning strategies further address domain shift and rare-class imbalance [22]. Gaps. Many studies optimize either PV-panel detection or generic building/roof extraction in isolation, with performance often tied to specific sensors, regions, or ground sampling distances. Cross-city transfer remains brittle; instance-level quality (shape/extent) can lag in cluttered urban scenes; and compute-heavy backbones hinder deployment at scale or on low-power devices. Our research addresses practical deployability with a lightweight attention U-Net that preserves boundaries while keeping inference efficient, but like most 2D-only detectors still struggles where occlusions and complex roof geometry dominate.

2.2 Pixel-to-ground scaling and geodesy considerations

Accurate rooftop area requires converting pixels to meters with awareness of map projection and zoom/resolution. A common approach in Web-Mercator maps uses the latitude- and zoom-dependent ground resolution (meters-per-pixel), which is only locally valid and degrades with latitude and off-nadir perspective [3, 16]. Production toolchains often avoid this pitfall by consuming georeferenced imagery/tiles or projecting roof polygons to equal-area systems before measuring [2]. Gaps. Many rooftop/PV studies gloss over scale provenance or error propagation from pixel resolution to area especially when harvesting non-georeferenced web tiles. In contrast, this research explicitly implements a pixel-to-ground factor for Google Static Maps and validates area against Google Earth measurements, reporting mean errors of 11% (web tiles) and 3% (dataset imagery), highlighting both feasibility and the need for careful scale handling.

2.3 Imagery-based rooftop PV assessment

At global scale, Joshi et al. combine machine learning and geospatial analytics to estimate 27 PWh-yr from rooftops, illustrating the policy value of consistent, high-resolution mapping [10]. City-scale methods increasingly fuse DL segmentation with 3D/DSM context to capture shading, tilt, and usable roof patches; Ren et al. integrate deep learning with a 3D-GIS irradiance analyzer and show that shading and availability jointly reduce annual energy by up to 36% in dense Hong Kong, cautioning against additive reductions [17]. Orientation-aware PV potential models and end-to-end DL+GIS frameworks further improve irradiance/production estimates when reliable DSMs/LiDAR are available [12, 13]. Project Sunroof demonstrates an industrial-scale pipeline using aerial photogrammetry DSMs to

compute per-pixel shading and technical potential [1]. Gaps. Many SOTA pipelines depend on expensive 3D data or limited-coverage DSMs; others extrapolate energy from 2D segmentation without explicit shading/tilt or empirical PV generation validation. Label scarcity and domain shift also limit generalization.

2.4 Positioning this study

This research addresses the above gaps through an integrated, data-light pipeline. The novelty lies in combining: (i) a Full-Attention U-Net that sharpens rooftop boundaries and reduces false positives, (ii) an explicit pixel-to-ground scaling procedure validated across georeferenced and web-map imagery, and (iii) the translation of rooftop masks into actionable planning-grade PV indicators such as installable capacity, indicative yield, and ward-level aggregates. Unlike prior work that optimises segmentation metrics alone, our framework demonstrates how attention-based architectures and careful geodesy can deliver robust outputs transferable across cities. Crucially, the approach remains scalable and policy-relevant. By avoiding reliance on LiDAR or dense cadastral data, the pipeline can be deployed in data-sparse contexts, supporting net-zero planning, public asset prioritisation, and feeder-level grid coordination. In this way, it offers a pragmatic bridge between computer vision innovation and the evidence-based decision-making required in urban sustainability practice.

3 Methodology

3.1 Dataset and Pre-processing

We train and evaluate on the INRIA Aerial Image Labeling Dataset: 5000×5000-px orthotiles at 0.3 m/px from 10 cities with building-footprint ground truth suited to rooftop delineation and downstream PV estimation [14]. From 200 tiles, we generate 384×384 crops with 30% overlap, yielding 38.9k image–mask pairs; splits follow a standard train/val/test protocol. Basic normalization and tiling are applied; labels retain binary rooftops for a single-class segmentation task.

3.2 Attention-Augmented Architectures

We compare three encoder–decoder variants built around U-Net. The baseline U-Net serves as a strong pixel-level reference with plain skip connections and a 1×1 sigmoid head for binary masks. The Skip-AG U-Net inserts attention gates (AGs) on every skip path, following [15]: decoder gating signals modulate encoder features so that only salient rooftop responses are forwarded at concatenation, improving edge sharpness and suppressing background clutter. Our Full-AG U-Net extends this idea by deploying AGs not only on skip paths but also within encoder and decoder stages, providing end-to-end saliency control across scales; empirically this yields the best boundaries and fewer false positives on complex roofs.

All models share a reference configuration: 384×384 inputs; encoder channels [64, 128, 256, 512] with two convolutions per stage; a 1024-channel bottleneck; and a symmetric decoder with up-convolutions and paired convolutions, culminating in a 1×1 convolution with sigmoid activation. We implement the networks in TensorFlow/Keras with modular blocks; AGs are realized as additive attention using gating from the decoder to reweight encoder activations before concatenation [15]. Training uses Adam optimization, binary

cross-entropy, early stopping, checkpointing, and learning-rate reduction on plateau. We report Accuracy, IoU, Dice, Precision/Recall, and ROC-AUC to capture both set-overlap and boundary-sensitive behaviour. In practice, the Full-AG variant provides the most reliable masks for downstream PV area estimation and layout simulation, while retaining inference efficiency suitable for city-scale tiling.

3.3 Pixel-to-Ground Scaling

We convert pixel areas to m² using a Web-Mercator ground-resolution factor derived from:

$$\mathrm{mpp}(z,\varphi) = \frac{156543.03392\cos\varphi}{2^z}$$

Width/height and total ground area follow from this adjusted resolution; a per-pixel conversion factor scales contour areas from masks. We validate areas against Google Earth measurements: mean errors 11.0% on Google Static Maps tiles versus 3.2% on georeferenced dataset crops, indicating planning-grade fidelity.

3.4 Obstruction Handling and Layout Constraints

Rooftop contours are extracted via OpenCV; oriented minimum-area rectangles provide facet orientation. Panels are raster-packed as 1.7 m×1.0 m rectangles with 0.5 m spacing, and each placement is validated to lie fully within the roof mask. North/south assignment uses dot-products with facet vectors for directional statistics. Current masks do not explicitly remove superstructures (e.g., vents, chimneys); shading, pitch and dynamic sun angles are not yet modeled which is a limitation to this study.

3.5 Energy-Modelling Stack

For fast scoping, we aggregate placed-panel counts to capacity and simple yield proxies. Per-section energy is estimated as:

$$\text{Energy (W)} = \text{PanelCount} \times 300 \times \text{DirectionFactor} \times 0.20$$

with UK-typical irradiance factors (south = 100%, north = 50%). Outputs include building-level capacity/yield and ward-level aggregates for planning. Segmentation metrics (e.g. IoU) are reported alongside area-error statistics to connect pixel accuracy with PV-relevant geometry quality.

4 Results

Across 38.9k test image–mask pairs derived from INRIA aerial tiles, the attention-augmented architectures consistently improved rooftop segmentation quality over a plain U-Net [14, 15]. The Full-AG U-Net achieved the highest test IoU (0.874) with strong validation accuracy (0.965) and low validation loss (0.095), outperforming both the baseline U-Net (IoU = 0.773) and the Skip-AG variant (IoU = 0.847). These gains align with the intuition that attention improves boundary fidelity and suppresses background clutter in dense urban scenes. The dataset scale and split protocol (38,988 crops; train/val/test) remained consistent to underpin comparability across runs.

To connect pixel-level accuracy to energy-relevant geometry, predicted masks were converted to geospatial areas via a tile-wise metres-per-pixel factor and validated against Google Earth measurements. Using Google Static Maps tiles, mean area error was 11.02%; using INRIA/georeferenced crops, mean area error dropped to 3.23%, indicating planning-grade fidelity when imagery is georeferenced and scale factors are applied consistently. These findings are supported by per-building comparisons showing typical absolute percentage errors in single digits for most roofs (e.g., 0.18–3.84%), with larger errors appearing on a minority of complex geometries. Qualitatively, attention-enabled models converged faster and exhibited reduced overfitting behaviour consistent with improved generalisation under varied roof shapes and textures.

The pipeline's energy estimation used deterministic rules (panel count × 300 W × direction factor × efficiency) to produce building-level capacity/yield proxies and aggregated indicators for planning (wards/LSOAs). Because sampled roofs had no operational PV systems, energy figures are predictive and presented as modeled potential only; empirical validation against smart-meter/SCADA data was not possible in this study. Still, the close agreement in area (a primary driver of capacity) suggests the outputs are suitable for preliminary siting, targeting of public assets, and scenario screening by local authorities and DSOs. Table 1 consolidates training/validation metrics and IoU for the three models as reported in this research, with Full-AG U-Net leading on all key indicators.

Table 1: Training/validation metrics and test IoU (as reported).

Model	Train Acc	Val Acc	Val Loss	IoU
U–Net (baseline)	0.944	0.937	0.170	0.773
Skip–AG U–Net	0.966	0.946	0.143	0.847
Full–AG U–Net	**0.977**	**0.965**	**0.095**	**0.874**

Table 2: Area estimation summary vs. Google Earth measurements.

Image Source	Computed (m^2)	Estimated (m^2)	Error (%)
Google Static Maps	1264.97	1119.64	11.02
INRIA (georeferenced)	1871.49	1851.92	3.23

Table 2 summarises the aggregate area-error comparison between web-tile and georeferenced imagery. Results favour georeferenced inputs, as expected from reduced scale/projection uncertainty. Figure 1 shows qualitative performance of the proposed model, using sample image from the dataset.

The results show that attention mechanisms materially improve rooftop segmentation quality and stability, translating into lower area error after pixel-to-ground conversion. The 3% mean error on georeferenced imagery is within tolerance for planning-grade PV screening, enabling credible capacity/yield mapping where detailed 3D data are unavailable. Remaining gaps stem from 2D assumptions (no explicit tilt/shading/superstructures) and the lack of measured PV output for calibration. Addressing these will likely require (i) integration of 3D geometry (LiDAR/photogrammetry) for facet tilt and shading; (ii) uncertainty quantification from segmentation through energy modelling; and (iii) validation against operational datasets to reconcile modeled and realised yield directions. Overall, the evidence supports a pragmatic conclusion: Full-AG U-Net provides the most reliable masks for downstream PV estimation; pixel-to-ground scaling is accurate when georeferencing is sound; and

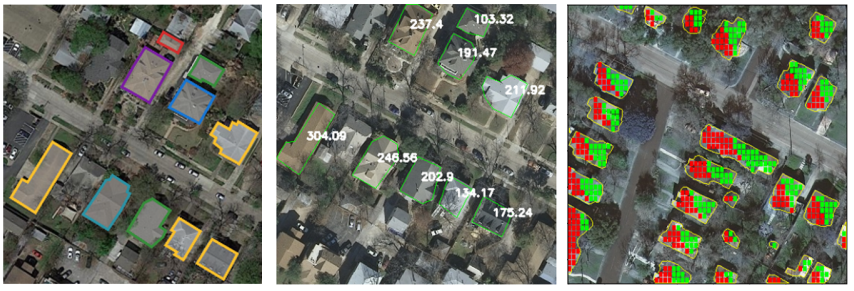

Figure 1: From left to right, Google Maps Computed Area, Estimated Area, and Panels Placement Simulation

the data-light energy proxy is fit for early-stage planning, pending calibration to measured generation in future work.

5 Discussion

This study introduced a data-light pipeline that combines attention-augmented U-Net segmentation, explicit pixel-to-ground scaling, and the translation of rooftop masks into planning-grade photovoltaic (PV) indicators. The key novelty lies in integrating these three components into a unified framework that moves beyond benchmark optimisation and delivers outputs that are directly useful to planners and distribution system operators. Unlike many prior approaches that emphasise segmentation accuracy in isolation, this pipeline demonstrates how deep learning architectures can be linked to geospatial scaling and energy modelling to produce decision-relevant insights. At the same time, several limitations should be acknowledged. The reliance on 2D segmentation restricts the capacity to model roof tilt, shading, and structural obstructions such as chimneys and dormers. These factors are critical in determining usable area and effective generation potential. Integrating 3D data sources such as LiDAR, photogrammetry, or DSMs would allow facet-based segmentation and shading-aware modelling, addressing one of the main sources of error in 2D-only pipelines. Related to this, current obstruction handling is basic and does not fully capture superstructures or dynamic shadows; targeted detection and shadow-casting approaches could improve fidelity. Another important limitation is that energy estimates remain theoretical. By deriving outputs from geometry and simple irradiance proxies, the framework ensures portability across contexts but lacks empirical grounding. Future work will require operational validation against real-world PV generation data (e.g., smart-meter or SCADA feeds). Such validation would allow calibration of model assumptions, reduce systematic biases, and demonstrate applicability at scale.

Uncertainty quantification also deserves explicit attention. At present, the pipeline produces deterministic indicators without error bounds. Incorporating probabilistic models or Bayesian deep learning approaches could provide confidence intervals on both segmentation outputs and energy estimates. This would improve transparency and give policymakers more reliable evidence for planning decisions. Scalability is a major strength of this approach, yet explicit testing across geographies with varied architectural typologies and climates is needed to demonstrate robustness. Comparative studies in different cities would clarify

adaptability and highlight contexts where retraining or fine-tuning is required. Finally, while the Full-Attention U-Net achieved strong IoU and accuracy, broader benchmarking against state-of-the-art models such as transformer-based or Mask2Former architectures, and reporting additional metrics such as precision, recall, and computational efficiency, would strengthen the evaluation.

Overall, while limitations remain, the pipeline already offers a pragmatic, lightweight, and transferable method that can support evidence-based planning. Its combination of accuracy, scalability, and policy-relevant outputs highlights its potential as a bridge between computer vision innovation and the urgent practical needs of urban sustainability.

6 Conclusion

This paper presented a data-light, end-to-end pipeline that converts aerial imagery into planning-grade indicators of rooftop PV potential. By pairing attention-augmented U-Net architectures with a careful pixel-to-ground scaling procedure, our approach delivers accurate rooftop masks and stable area estimates that can be translated into installable capacity and indicative yields. In contrast to many state-of-the-art studies that optimise segmentation metrics without operational outputs, we emphasised products that cities and distribution system operators can act on building-level suitability, aggregated capacity/yield by ward, and simple carbon-abatement signals to support targeting of public assets and energy-poverty interventions. Empirically, attention mechanisms improved boundary fidelity and reduced false positives, and georeferenced imagery enabled low area error suitable for early-stage planning. While our energy calculations are intentionally simple to remain portable across locales, they provide a credible first pass where detailed 3D data or measured PV generation are unavailable.

A central strength of the proposed framework is its portability and scalability. Unlike pipelines that depend on dense cadastral data, high-resolution LiDAR, or bespoke city-specific calibration, our approach is deliberately data-light. This ensures that rooftop PV assessments can be generated even in data-sparse contexts where municipal datasets are incomplete or inaccessible. The ability to scale rapidly across jurisdictions without major preprocessing makes the pipeline suitable for cross-city benchmarking, longitudinal monitoring, and integration into national net-zero planning initiatives. Beyond technical accuracy, the broader impact lies in bridging computer vision with urban sustainability practice. The outputs of this pipeline are designed in the language of policy capacity maps, indicative yields, feeder-level demand offsets, and asset-prioritisation metrics rather than isolated computer vision scores. In this way, our study provides a pragmatic bridge between research advances and actionable planning insights. This alignment with decision-making needs strengthens the role of AI-powered geospatial analytics in supporting energy transitions, carbon reduction strategies, and evidence-based policy interventions.

The principal limitations are the current reliance on 2D geometry (no explicit tilt, shading, or superstructure masking) and the absence of validation against operational PV output. Future work will integrate DSM/LiDAR-derived roof facets and shading, incorporate uncertainty propagation from segmentation through energy modelling, and validate against smart-meter or SCADA data. We also see value in domain-adaptation strategies for cross-city transfer, temporal analyses to capture roof changes, and coupling with socio-economic adoption models. Overall, the proposed pipeline offers a pragmatic bridge between computer vision advances and urban sustainability practice enabling scalable, comparable, and

decision-relevant rooftop PV assessments that support net-zero planning and grid coordination.

References

[1] Project sunroof data explorer: Methodology. Technical report, Google, 2018. URL https://www.google.com/get/sunroof/assets/data-explorer-methodology.pdf. Accessed 18 Aug 2025.

[2] Zoom levels and scale. Esri Developer Documentation, 2025. URL https://developers.arcgis.com/documentation/mapping-and-location-services/reference/zoom-levels-and-scale/. Accessed 18 Aug 2025.

[3] Zoom levels. https://wiki.openstreetmap.org/wiki/Zoom_levels, 2025. Accessed 18 Aug 2025.

[4] Jeff Anderson, Scott Earl, Kat Bristol, et al. Project sunroof methodology. Google, 2018. URL https://static.googleusercontent.com/media/sunroof.withgoogle.com/en//static/sunroof-methodology.pdf.

[5] Katalin Bódis, Ioannis Kougias, Arnulf Jäger-Waldau, Nick Taylor, and Sándor Szabó. A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the european union. *Renewable and Sustainable Energy Reviews*, 114:109309, 2019. doi: 10.1016/j.rser.2019.109309.

[6] Climate Action Network (CAN) Europe. Rooftop solar pv: Country comparison report—update 2024. Technical report, 2024. URL https://caneurope.org/content/uploads/2024/04/Rooftop-Solar-PV-Report-Update_April-2024.pdf.

[7] Aron P. Dobos. Pvwatts version 5 manual. Technical Report NREL/TP-6A20-62641, National Renewable Energy Laboratory (NREL), Golden, CO, 2014. URL https://docs.nrel.gov/docs/fy14osti/62641.pdf.

[8] G. García et al. Generalized deep learning model for photovoltaic module segmentation from satellite and aerial imagery. *Solar Energy*, 274:112539, 2024. URL https://www.sciencedirect.com/science/article/abs/pii/S0038092X24002330. Mask2Former-based PV segmentation.

[9] Thomas Huld, Richard Mueller, and Attilio Gambardella. A new solar radiation database for estimating pv performance in europe and africa. *Solar Energy*, 86(6): 1803–1815, 2012. doi: 10.1016/j.solener.2012.03.006.

[10] Shivika Joshi et al. High-resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation. *Nature Communications*, 12(5738), 2021. doi: 10.1038/s41467-021-25720-2.

[11] Siddharth Joshi, Shivika Mittal, Paul Holloway, Priyadarshi R. Shukla, Brian Ó Gallachoir, and James Glynn. High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation. *Nature Communications*, 12(5738), 2021. doi: 10.1038/s41467-021-25720-2.

[12] Q. Li, S. Krapf, L. Mou, Y. Shi, and X. X. Zhu. Deep learning-based framework for city-scale rooftop solar potential estimation by considering roof superstructures. *Applied Energy*, 374:123456, 2024. URL https://ideas.repec.org/a/eee/appene/v359y2024ics030626192400103x.html. Online first; details per publisher.

[13] W. Lin et al. A robust deep learning framework for solar potential estimation with orientation-invariant rooftop classification. In *ISPRS Archives*, 2024. URL https://isprs-archives.copernicus.org/articles/XLVIII-1-2024/371/2024/isprs-archives-XLVIII-1-2024-371-2024.pdf.

[14] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3226–3229, 2017. doi: 10.1109/IGARSS.2017.8127684.

[15] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. arXiv:1804.03999, 2018. URL https://arxiv.org/abs/1804.03999.

[16] W. Reich and M. Hässig. Some principles of web mercator maps and their computation. Technical report, Universität der Bundeswehr München, 2020. URL https://athene-forschung.unibw.de/doc/132233/132233.pdf.

[17] Haoshan Ren, Chengliang Xu, Zhenjun Ma, and Yongjun Sun. A novel 3d-gis and deep learning integrated approach for high-accuracy rooftop solar energy potential characterization of high-density cities. *Applied Energy*, 306:117985, 2022. doi: 10.1016/j.apenergy.2021.117985.

[18] U.S. Department of Energy. Solar power in your community guidebook. Technical report, U.S. Department of Energy, Washington, DC, 2023. URL https://www.energy.gov/sites/default/files/2023-03/Solar_Power_in_Your_Community_Guidebook_March2023.pdf.

[19] J. Wang, X. Chen, W. Shi, W. Jiang, X. Zhang, H. Li, J. Liu, and H. Sui. Rooftop pv segmenter: A size-aware network for segmenting rooftop photovoltaic systems from high-resolution imagery. *Remote Sensing*, 15(21):5232, 2023. doi: 10.3390/rs15215232.

[20] Jianxun Wang, Xin Chen, Weiyue Shi, Weicheng Jiang, Xiaopu Zhang, Li Hua, Junyi Liu, and Haigang Sui. Rooftop pv segmenter: A size-aware network for segmenting rooftop photovoltaic systems from high-resolution imagery. *Remote Sensing*, 15(21):5232, 2023. doi: 10.3390/rs15215232.

[21] Jiafan Yu, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states. *Joule*, 2:2605–2617, 2018. doi: 10.1016/j.joule.2018.11.021.

[22] M. Zech et al. Toward global rooftop pv detection with deep active learning. *Energy and AI*, 2024. URL https://www.sciencedirect.com/science/article/pii/S2666792424000295. In press/online.

[23] Teng Zhong, Zhixin Zhang, Min Chen, Kai Zhang, Zixuan Zhou, Rui Zhu, Yijie Wang, Guonian Lü, and Jinyue Yan. A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy*, 298:117132, 2021. doi: 10.1016/j.apenergy.2021.117132.