DEEPSOLAR-3M: AN AI-ENABLED SOLAR PV DATABASE DOCUMENTING 3 MILLION SYSTEMS ACROSS THE US

Rajanie Prabha, Zhecheng Wang, Chad Zanocco, June Flora, Ram Rajagopal
Civil and Environmental Engineering
Stanford University
Stanford, CA 94305
{rajanie,zhecheng,czanocco,jflora,ramr}@stanford.edu

ABSTRACT

The widespread deployment of distributed energy resources (DERs), especially solar photovoltaic (PV) systems, is essential for attaining a sustainable energy future. Uneven solar adoption slows decarbonization and entrenches energy poverty in vulnerable communities. ML-driven spatial analytics can pinpoint these gaps to prioritize policy interventions. Granular spatial mapping of all installed PV systems provides comprehensive data to facilitate a more equitable and efficient transition to sustainable energy sources. Previously, DeepSolar provided the most comprehensive nationwide solar PV dataset in the United States; however, it only extends up to mid-2017. This paper introduces a novel pipeline leveraging vision transformer models to detect rooftop-mounted solar PV systems at the building level, extending coverage through 2022. Our findings indicate that rooftop PV systems in the U.S. have doubled over the past five years, totaling approximately 2.95 million systems, with increased adoption observed across all states. The final dataset[1], which we are making publicly accessible, serves as an invaluable resource for policymakers, developers, researchers, and utilities dedicated to advancing equitable decarbonization efforts nationwide.

1 INTRODUCTION

Over the last decade, the share of electricity generated by renewable energy sources in the U.S. increased by 10%, reflecting a broader shift towards sustainable energy systems. Solar energy, in particular, has been a key driver of this growth. According to the National Renewable Energy Laboratory (NREL) Feldman et al. (2023), in 2022 alone, the United States added 17.0 GWac of solar PV capacity, with residential installations contributing 5.2 GWac—an increase of 30% compared to 2021. However, solar adoption is uneven, with certain states, counties, and block groups experiencing significantly faster growth than others. The disparity between solar adopters and non-adopters has significant distributional consequences, as those who adopt solar PV are typically wealthier than the general population (Barbose et al. (2018); Yu et al. (2018); Konzen et al. (2024)). Several factors are responsible for such widespread but uneven adoption such as falling costs of solar panels, high electricity prices, financial incentives, decarbonization goals, etc.

To develop policies and incentives that enhance equitable adoption of PV systems, it is essential to have current and detailed mapping data of these systems. Historically, conventional data collection methods such as surveys have been utilized; however, these methods often lack comprehensive coverage and spatial precision. The Berkeley Lab's "Tracking the Sun" Barbose et al. (2024) predominantly collects data from state agencies and utilities that manage PV incentive programs, maintain solar renewable energy credit registration systems, and supervise interconnection processes. To thoroughly understand equity dynamics, including off-grid systems is crucial. However, this information was gathered from 72 entities across just 31 states and does not account for off-grid PV systems, which are essential for comprehensively understanding equity dynamics. Recent research Kruitwa-

[1]Dataset link: https://github.com/rajanieprabha/DeepSolar-3M.git

gen et al. (2021) has increasingly employed remote sensing and aerial imagery, aided by computer vision algorithms, to observe PV systems at a large scale. The DeepSolar project Yu et al. (2018) employs deep learning models with high-resolution aerial imagery to identify PV systems across the United States, offering valuable insights into deployment trends, socio-economic equity gaps, and policy implications. Although the DeepSolar dataset significantly contributed to our understanding of solar adoption patterns up to 2017, the recent rapid expansion of solar installations has rendered the original dataset outdated and insufficient for current analysis needs.

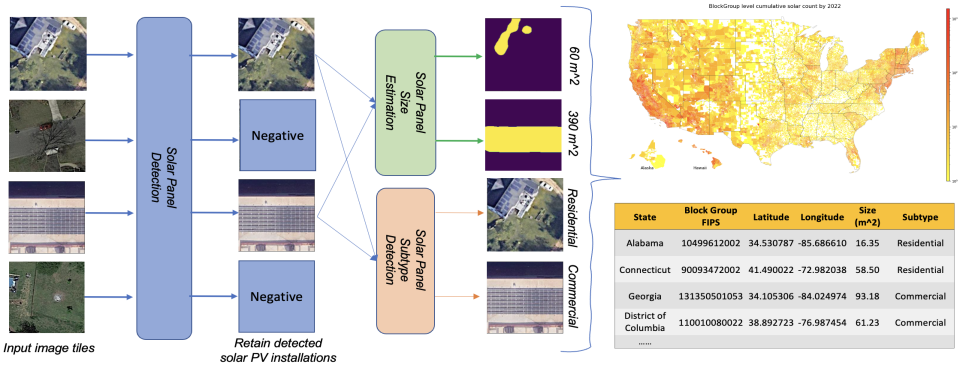


Figure 1: Solar PV Data Acquisition Framework

Aerial imagery is a valuable resource for detecting solar PV systems. However, its variability in terms of zoom level, resolution, and image quality requires careful handling. While computer vision models like convolutional neural networks (CNNs) are employed in this work, they inherently require large amounts of training data to effectively manage these challenges and achieve accurate results. Vision Transformers (ViTs) Dosovitskiy et al. (2021) offer a more adaptable approach by using self-attention mechanisms to capture long-range dependencies and contextual relationships across images. This adaptability is essential for integrating new data that may slightly differ in quality from the initial training set. We fine-tuned the pre-trained ViT model with various strategies (He et al. (2022)) to optimize our training dataset size and computational complexity.

There are two main contributions of this paper:

1. Our data acquisition pipeline as shown in Figure 1 has three models: PV detection binary classification model, PV semantic segmentation model, and PV type multiclass classification.
2. The resulting dataset 'DeepSolar-3M' encompasses approximately 2.95 million rooftop PV systems distributed across all 50 states and the District of Columbia with system-level information, such as location, size, and type (residential, commercial, or utility-scale).

To the best of our knowledge, this is, so far, the most complete and comprehensive U.S.-wide residential PV dataset. By integrating this latest data into energy policy frameworks, we can more effectively identify the adoption barriers in regions with lower adoption rates, thereby accelerating solar deployment in underserved communities.

2 Data Acquisition Pipeline

Data Collection. High-resolution imagery is needed to accurately detect solar PV systems, especially residential systems. To reduce the geographic scope to scan, we focused only on buildings as the majority of non-utility scale PV systems are installed on rooftops. The Microsoft Maps dataset Microsoft (2023) contains 129,591,852 computer-generated building footprints, spanning all of the US states and the District of Columbia, derived using computer vision algorithms on satellite imagery. Using this dataset, we collected around 230 million image tiles with a resolution of less than 30 cm per pixel using the Google Maps khms API. Each image tile consists of three channels, with dimensions of 256 by 256 pixels, and is acquired at a zoom level of 20. A few sample images are

shown in Appendix figure 4. Furthermore, due to the significant variation in building density across the country, certain states have a substantially larger number of image tiles compared to others.

Pipeline Set-up. We followed a three-phase pipeline naming — Detection, Segmentation, and Type-Classification. Detection constitutes the initial process, wherein a binary label is assigned to determine the presence of PV pixels within an image tile. Once all images identified as lacking PV pixels (negative cases) are filtered out, the remaining images proceed to the second phase of analysis. In this phase, Segmentation and Type-Classification are conducted concurrently. The Segmentation model provides information regarding the size of the PV system, while the Type-Classification model categorizes the PV system into one of the following classifications: residential PV, commercial PV, utility-scale PV, or solar heating system. For each task in the pipeline, we employed distinct supervised computer vision models to ensure accuracy and precision. However, this method requires extensive labeling and is computationally intensive. To address these challenges, we explored various fine-tuning strategies for training the binary classification PV detection model. Specifically, we fine-tuned a Vision Transformer Masked Autoencoders (ViTMAE) He et al. (2021) model, self-supervised pre-trained on the ImageNet-1k (IN1K) Deng et al. (2009) dataset, using four different fine-tuning strategies (see Appendix A.3.1 for details).

Detection. We created the fine-tuning dataset consisting of only 45,000 training and 5,000 validation labels to train the binary classification model. Even though this training set has 85% fewer training labels than the CNN-based PV detection model in DeepSolar Yu et al. (2018), it is adequate to achieve similar performance with the *LoRA* fine-tuning approach. Our *LoRA* fine-tuning reduces compute costs by 99% versus CNNs, enabling frequent nationwide updates critical to tracking equity progress and sustainable development goals. The training was performed for 20 epochs with an Adam optimizer and a learning rate 10^{-6} following a linear schedule and L2 regularization weight decay 1.0 to minimize any over-fitting. Subsequently, we assessed our model's performance using an in-house evaluation dataset with $92,000$ image labels out of which $3,455$ are positive PV labels spread across the U.S. Our best model's test set precision is 0.94 and recall is 0.91. The final model was deployed on all the image tiles collected to get positive detections across the U.S.

Segmentation. In the pipeline's second phase, we used another model called Segformer Xie et al. (2021) to get the solar PV segmentation boundary from the image tile. The Segformer model, also pre-trained on the IM1K dataset, was finetuned with $5,607$ supervised PV training labels for the semantic segmentation task. The model was further validated on 300 labels. The reported mean Intersection Over Union (IoU) on the test set comprising 600 labels is 0.92, while the IoU for the PV class specifically is 0.86. A few prediction results are shown in Appendix A.4. After deploying this model on all the positive PV image tiles, we got the segmentation mask for each image tile. The final PV size in square meters (m^2) was calculated by converting the image pixels into real-world distances.

Type-Classification. The final phase involves merging the image tiles so that a single solar PV system split across multiple image tiles can be combined as one solar system. Simultaneously, we also classify the PV systems into various categories: residential PV, commercial PV, utility-scale PV, and solar heating system. For subtype classification, the Resnet 50 Szegedy et al. (2015) model was finetuned on $10,000$ supervised image training labels and tested on 1,179 image labels to predict the type of solar PV system (results are shown in Appendix A.5).

3 RESULTS

Our ML pipeline identified 2.95 million rooftop solar PV systems throughout the US, depicting 142 km^2 of panel area, with 2.5 million classified as residential. This marks a 250% increase in residential installations since 2017. As shown in Figure 2 (left), California remains the leader in total installations (43% of the national share in 2022), although its dominance has decreased from 57% in 2017. Texas saw the most significant growth, doubling its share of residential installations (2% to 4%), while Arizona's contribution decreased by 2%.

Using irradiance data from NASA Power API, residential PV capacities were estimated (Fig. 2, right). The data reveals substantial state-level disparities, with averages ranging from 2.5 kW_{dc}(Alaska) to 8.1 kW_{dc}(Arizona). Right-skewed distributions (averages $>$ medians in all states) suggest a subset of households install disproportionately larger systems, likely driven by income in-

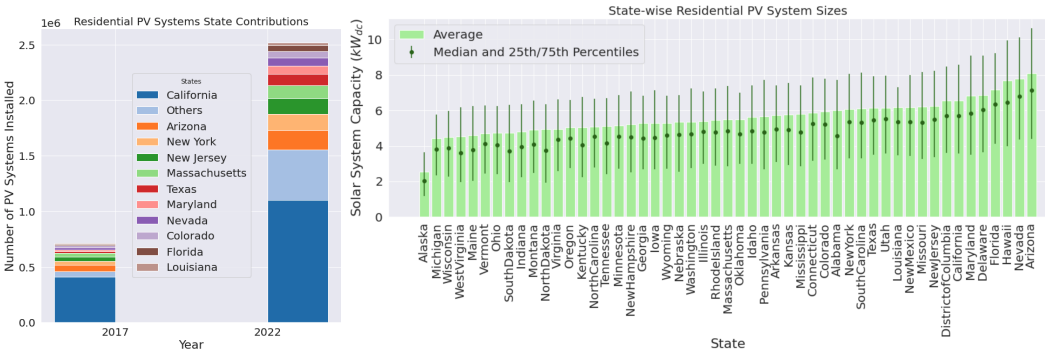


Figure 2: (left) Bar plot comparing Residential PV Systems Count in 2017 and 2022. (right) State-wise Residential PV System Capacities with Average, Median, and 25th/75th Percentiles.

equality or incentives favoring high adopters. For example, in Nevada and Arizona —states with the highest averages —the 75th percentiles are around 10 kW_{dc}, indicating conditions that encourage larger installations. Conversely, lower-capacity states (average <5 kW_{dc} may reflect affordability barriers or less favorable incentive structures. These disparities underscore the need for equity-focused policies, such as tiered rebates for smaller systems in underserved regions, to broaden participation and mitigate adoption gaps.

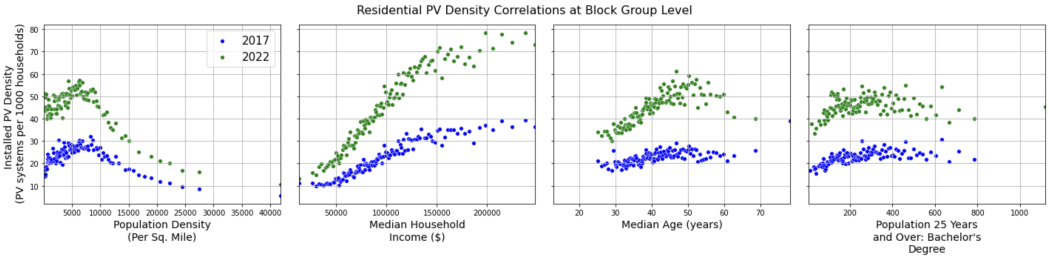


Figure 3: PV density (PV systems per 1,000 housing units) has increased from 2017 to 2022 across all groups. PV density peaks at a population density of ∼7,000 capita per square mile, decreasing in more urbanized areas. Higher-income areas show a clear upward trend in PV density, plateauing around $150k in 2017. Older populations and those with higher education levels also exhibit greater PV densities, highlighting the significant influence of demographic and socioeconomic factors on solar PV adoption

4 Conclusion

The rapid growth and geographic diversification of residential solar adoption underscore its potential to democratize renewable energy access. However, persistent equity gaps—evident in the correlation between PV density and various socioeconomics (Figure 3)—highlight systemic barriers. Our AI-enabled pipeline—demonstrates machine learning's transformative role in mapping and accelerating equitable renewable energy transitions. A few critical insights emerge from our analysis, underpinned by AI's capacity to process continental-scale imagery and socio-economic data:

1. Geographic adoption of residential solar expanded significantly between 2017–2022: 50% of census block groups accounted for 90% of PV installations in 2022 (up from 20% in 2017), with over 40% of block groups installing their first residential PV system during this period, reflecting penetration into new regions.
2. Our data identifies $\sim$ 20,000 block groups with high solar potential (irradiance $>$ 5 kWh/m^2/day) but adoption rates, defined by % of housing units in a block group with residential PV, less than 1%—priority zones for community solar programs.

3. ML reveals 45% of zero-adoption areas (block groups with no residential PVs) have >40% renters. Targeted virtual net metering could unlock 18 GW here—equivalent to powering 12 million homes, assuming no policy rollout delays. (see appendix A.6 for details).

Our work exemplifies how AI accelerates climate solutions by converting raw satellite data into policy-ready insights. As climate deadlines loom, AI-enabled databases like DeepSolar-3M provide the spatial intelligence needed to prioritize investments that maximize emissions reductions and social impact. By democratizing access to granular adoption analytics, AI becomes not just a technical tool but a bridge to equitable decarbonization.

REFERENCES

G. L. Barbose, N. R. Darghouth, B. Hoen, and R. H. Wiser. Income Trends of Residential PV Adopters: An analysis of household-level income estimates. April 2018. URL https://escholarship.org/uc/item/7p21s3nz.

Galen Barbose, Naïm Darghouth, Eric O'Shaughnessy, and Sydney Forrester. Tracking the Sun: Pricing and Design Trends for Distributed Photovoltaic Systems in the United States, 2024 Edition. Technical report, Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States), August 2024. URL https://www.osti.gov/biblio/2438480.

Xinlei Chen, Saining Xie, and Kaiming He. An Empirical Study of Training Self-Supervised Vision Transformers, August 2021. URL http://arxiv.org/abs/2104.02057. arXiv:2104.02057 [cs].

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848. URL https://ieeexplore.ieee.org/document/5206848. ISSN: 1063-6919.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, June 2021. URL http://arxiv.org/abs/2010.11929. arXiv:2010.11929 [cs].

David Feldman, Krysta Dummit, Jarett Zuboy, and Robert Margolis. Spring 2023 Solar Industry Update [Slides]. Technical Report NREL/PR–7A40-86215, 1974994, MainId:86988, April 2023. URL https://www.osti.gov/servlets/purl/1974994/.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked Autoencoders Are Scalable Vision Learners, November 2021. URL https://arxiv.org/abs/2111.06377v3.

Yutong He, Dingjie Wang, Nicholas Lai, William Zhang, Chenlin Meng, Marshall Burke, David B. Lobell, and Stefano Ermon. Spatial-Temporal Super-Resolution of Satellite Imagery via Conditional Pixel Synthesis, April 2022. URL http://arxiv.org/abs/2106.11485. arXiv:2106.11485 [cs].

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, October 2021. URL http://arxiv.org/abs/2106.09685. arXiv:2106.09685 [cs].

Gabriel Konzen, Rohan Best, and Nivalde José de Castro. The energy injustice of household solar energy: A systematic review of distributional disparities in residential rooftop solar adoption. *Energy Research & Social Science*, 111:103473, May 2024. ISSN 2214-6296. doi: 10.1016/j.erss.2024.103473. URL https://www.sciencedirect.com/science/article/pii/S2214629624000641.

L. Kruitwagen, K. T. Story, J. Friedrich, L. Byers, S. Skillman, and C. Hepburn. A global inventory of photovoltaic solar energy generating units. *Nature*, 598(7882):604–610, October 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03957-7. URL https://www.nature.com/articles/s41586-021-03957-7. Number: 7882 Publisher: Nature Publishing Group.

Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. Fine-Tuning can Distort Pretrained Features and Underperform Out-of-Distribution, February 2022. URL http://arxiv.org/abs/2202.10054. arXiv:2202.10054 [cs].

Microsoft. Building Footprints - Bing Maps, 2023. URL https://www.microsoft.com/en-us/maps/bing-maps/building-footprints.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015. IEEE. ISBN 978-1-4673-6964-0. doi: 10.1109/CVPR.2015.7298594. URL http://ieeexplore.ieee.org/document/7298594/.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, October 2021. URL http://arxiv.org/abs/2105.15203. arXiv:2105.15203 [cs].

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(12):2605–2617, December 2018. ISSN 2542-4351. doi: 10.1016/j.joule.2018.11.021. URL https://www.sciencedirect.com/science/article/pii/S2542435118305701.

A Appendix

A.1 Data

A few image tiles are shown in figure 4 and the image tile count state-wise distribution is shown in figure 5. California and Texas have the highest number of image tiles i.e., 11 million and 10 million buildings, respectively.

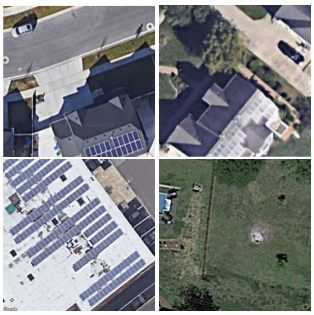


Figure 4: Samples of image tiles collected. Top row: Residential PVs, Bottom left: Commercial PVs, Bottom right: Negative.

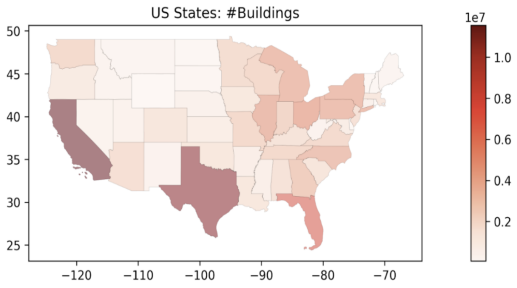


Figure 5: Image tile count distribution across the US

A.2 Evaluation Metrics

1. **F1 Score** is the harmonic mean of precision and recall. It provides a single metric that balances the trade-off between precision and recall, especially useful when the class distribution is imbalanced.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2. **Intersection over Union (IoU)** is a metric used to evaluate the accuracy of an object detector on a particular dataset. It measures the overlap between the predicted bounding box and the ground truth bounding box, defined as the area of their intersection divided by the area of their union.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$

A.3 Detection Model

The fine-tuning strategies explored as shown in figure 6 are:

A.3.1 Fine-Tuning Strategies

1. **Full network retraining** (*FT*): The most intuitive way of fine-tuning is retraining all the layers of the model with the downstream dataset. We used this as a benchmark to compare other fine-tuning strategies.
2. **Layer freezing** (L_k): Freezing the first or last layers has become a common practice for fine-tuning vision models and other language tasks now. In our case, we froze everything but the first two transformer blocks (L_2) for one set of experiments and similarly for the last two blocks (L_n).
3. **Linear probing** (*LP*): Following some of the literature Kumar et al. (2022); Chen et al. (2021) that posits larger gains from linear probing over fine-tuning in the presence of distribution shifts, we fine-tuned only the last MLP linear layer of each of the attention heads.

4. **Low-rank Matrix Factorization** (*LoRA*): The above fine-tuning methods still pose computational challenges (training parameters in millions). We used *LoRA* Hu et al. (2021), a technique that leverages the concept of *low intrinsic rank* to efficiently adapt pre-trained neural network models to new tasks. The idea behind LoRA is to constrain the weight updates by representing them with a low-rank decomposition. Specifically, for a pre-trained weight matrix $W_0 \in \mathbb{R}^{d\times k}$, the update is formulated as $W_0 + \Delta W = W_0 + BA$, where $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$ with $r \ll \min(d, k)$, ensuring that the model is both computationally efficient (less than 1% trainable parameters as compared with *FT)* and effective in adaptation. We used rank 4 for our experiments.

The F1 scores for all fine-tuning strategies on the eval set are shown in table 1. The evaluation data was labeled in-house and consists of 3,455 positive PV labels and 89,345 negative labels. A few true positives are shown in figure 7.

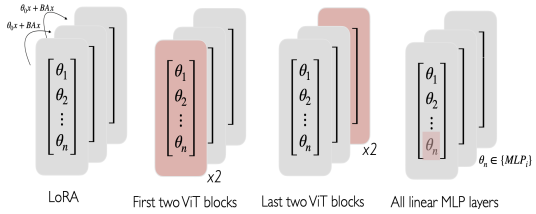


Figure 6: Fine-tuning strategies (left to right): Low-rank Matrix Factorization (LoRA), fine-tune only the first two transformer blocks (L_2), fine-tune only the last two transformer blocks (L_n), fine-tune all linear layers (LP)

Fine-tuning Strategy	F1-Score
FT	87.4
L_2	75.4
L_n	84.8
LP	85.8
LoRA	**92.4**

Table 1: Fine-tuning Results for the detection model

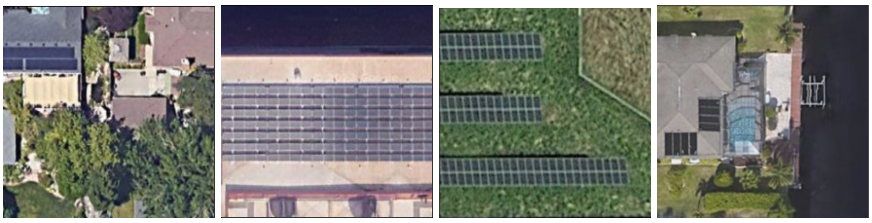


Figure 7: True Positives from the Detection Model

A.4 Segmentation Model

A few segmentation predictions for the eval set images are shown in figure 8. The whole eval set (600 labels) sample errors are the shown in figure 9.

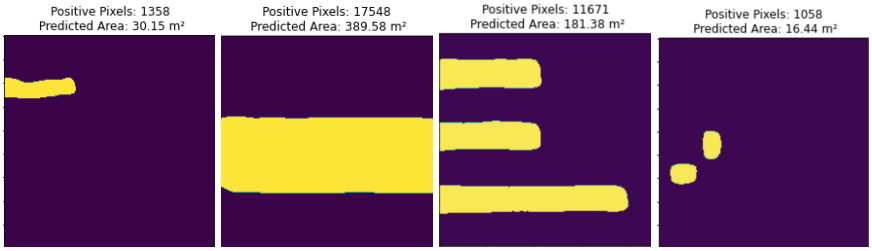


Figure 8: Segformer Model Predictions. The predictions are shown in terms of binary masks where the yellow color refers to the positively detected PV pixels which are then converted into m^2

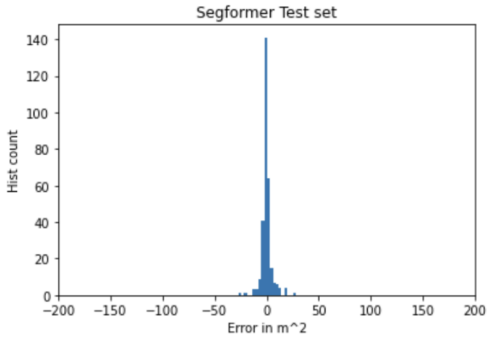


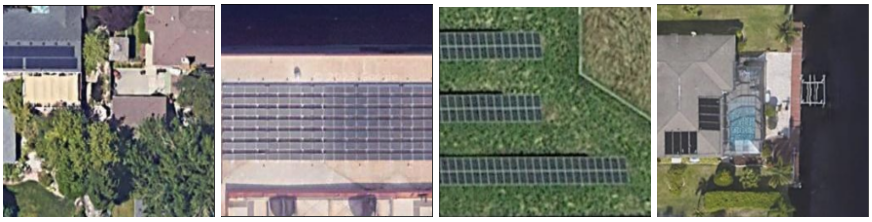
Figure 9: Segformer model evaluation set error (difference between ground truth and predicted labels in m^2) histogram

A.5 Type-Classification Model

The final model reported the following results on the evaluation set with 1,179 images (labeled in-house), shown in table 2. The diagonal values represent true positive cases, a few of those are shown in figure 10.

	Predicted				
Actual	**negative**	**solar-heat**	**residential**	**commercial**	**utility**
negative	**269**	10	13	5	0
solar-heat	7	**80**	7	0	0
residential	18	6	**670**	14	0
commercial	3	0	8	**50**	2
utility	0	0	0	2	**15**

Table 2: Confusion Matrix on Eval set for sub-type classification



Residential PV Commercial PV Utility Scale PV Solar Heating System

Figure 10: True Positives from the Subtype Classification model

A.6 Virtual Net-Metering Potential

Assuming 30% participation in rental-heavy areas:

$$P = 0.3 \times \mathbf{H}_{rental} \times \bar{P}_{system}$$

where $\mathbf{H}_{rental}$ refers to the number of housing units occupied by renters and $\bar{P}_{system}$ refers to the median capacity of a residential solar system.

Our dataset identified $\sim$ 30% of the block groups with no residential adoptions. Out of these block groups, $\sim$ 45% of them have more than $\sim$40% renter's occupied housing units (i.e. around 12 million renter's occupied housing units). Assuming $\bar{P}_{system}$ to be 5 kW_{dc}, we get:

$$P = 0.3 \times 12 \times 5 = 18\text{GW}$$