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
{rajanie,zhecheng, czanocco, jflora, ramr}@stanford.edu

Motivation

The widespread deployment of distributed energy resources (DERs), especially solar photovoltaic (PV) systems, is essential for a sustainable energy future. Uneven solar adoption slows decarbonization and deepens energy poverty in vulnerable communities. Granular spatial mapping of all installed PV systems enables a more equitable and efficient energy transition. While DeepSolar[1] previously provided the most comprehensive U.S. PV dataset, it only extends to mid-2017. This paper presents a new pipeline using vision transformer models to detect rooftop PV systems at the building level, extending coverage through 2022. Our findings show rooftop PVs in the U.S. have doubled in five years, reaching ~3 million systems, with growth across all states. The final dataset, released publicly, supports policymakers, developers, researchers, and utilities in advancing equitable decarbonization.

Dataset Benchmarking

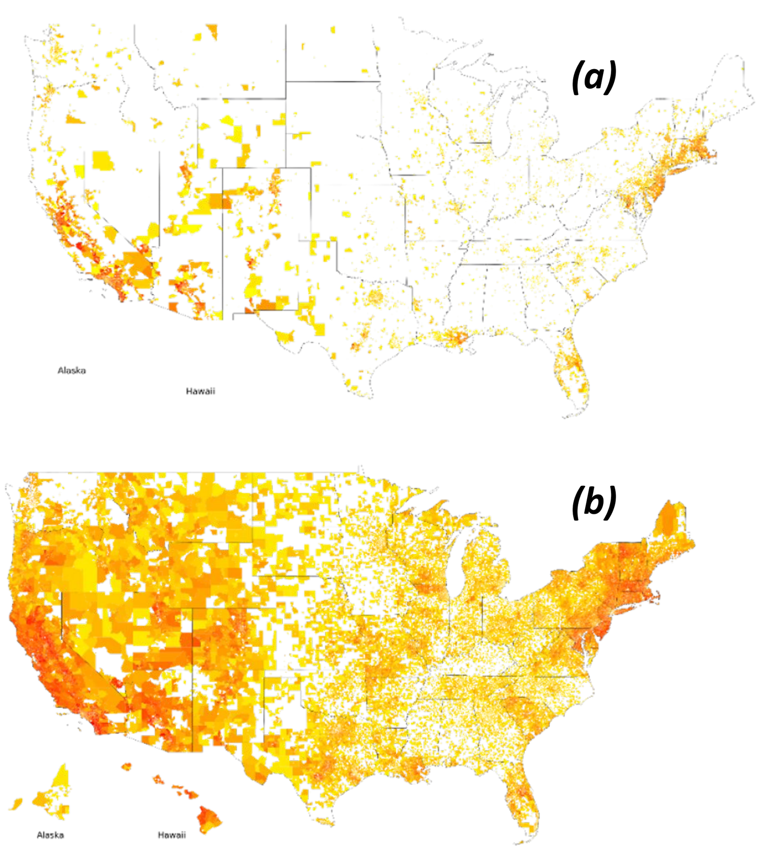

Figure 1. Nationwide Residential PV Dataset (a) 2017 (b) 2022

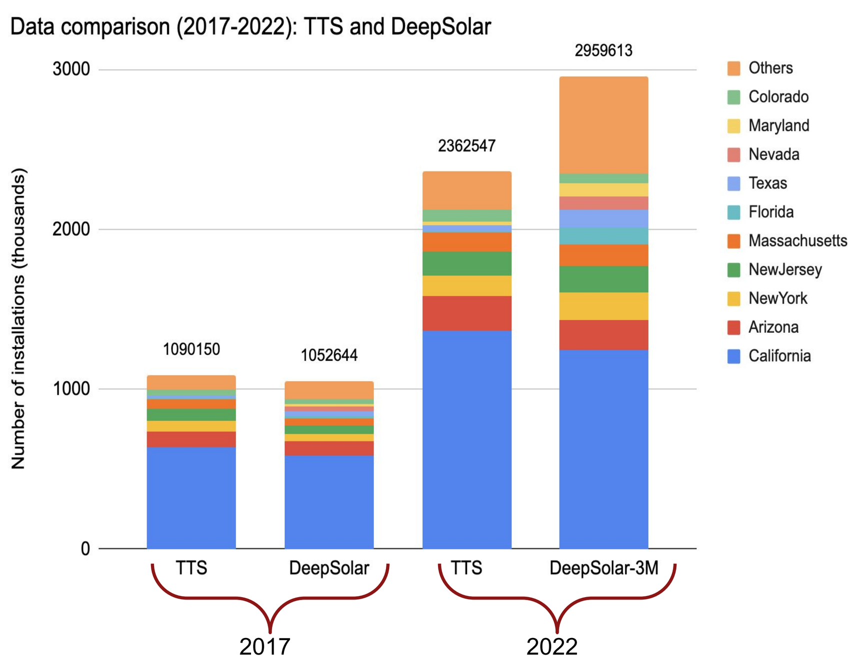

Figure 2. Data Benchmarking against Tracking the Sun[2]

- We identified nearly **3 million rooftop solar PV systems** using a fast-mapping AI pipeline, creating the most comprehensive nationwide PV map to date.
- Rooftop solar adoption has **nearly doubled since 2017**, with our model detecting ~600k more systems than the TTS dataset—highlighting both the power of machine learning and its critical role in addressing energy equity.

Data Acquisition Framework

To accurately detect rooftop solar PV systems, we used Microsoft's building footprint dataset to target buildings and collected ~230 million high-resolution (sub-30 cm) satellite image tiles across the U.S. via the Google Maps API. Using ViTMAE with LoRA fine-tuning, our detection model achieves 2017 DeepSolar-level PV detection accuracy while using 85% fewer labels and 99% less compute—enabling scalable, low-cost nationwide updates [F1 score: 92.4]. Subsequently, we fine-tuned a Segformer model to segment PV boundaries, achieving a mean IoU of 0.92. In the final phase, we used a fine-tuned ResNet-50 to classify PV types.

Figure 3. True Positives from the Detection Model

Residential, 30.15 m^2 | Commercial, 389.58 m^2 | Utility Scale, 181.38 m^2 | Solar Heater, 16.44 m^2

Figure 4. Segmentation & type-classification results

Dataset Overview

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. The data reveals substantial state-level disparities, with averages ranging from 2.5 kWdc(Alaska) to 8.1 kWdc(Arizona). Right-skewed distributions (averages > medians in all states) suggest a subset of households install disproportionately larger systems, likely driven by income inequality or incentives favoring high adopters.

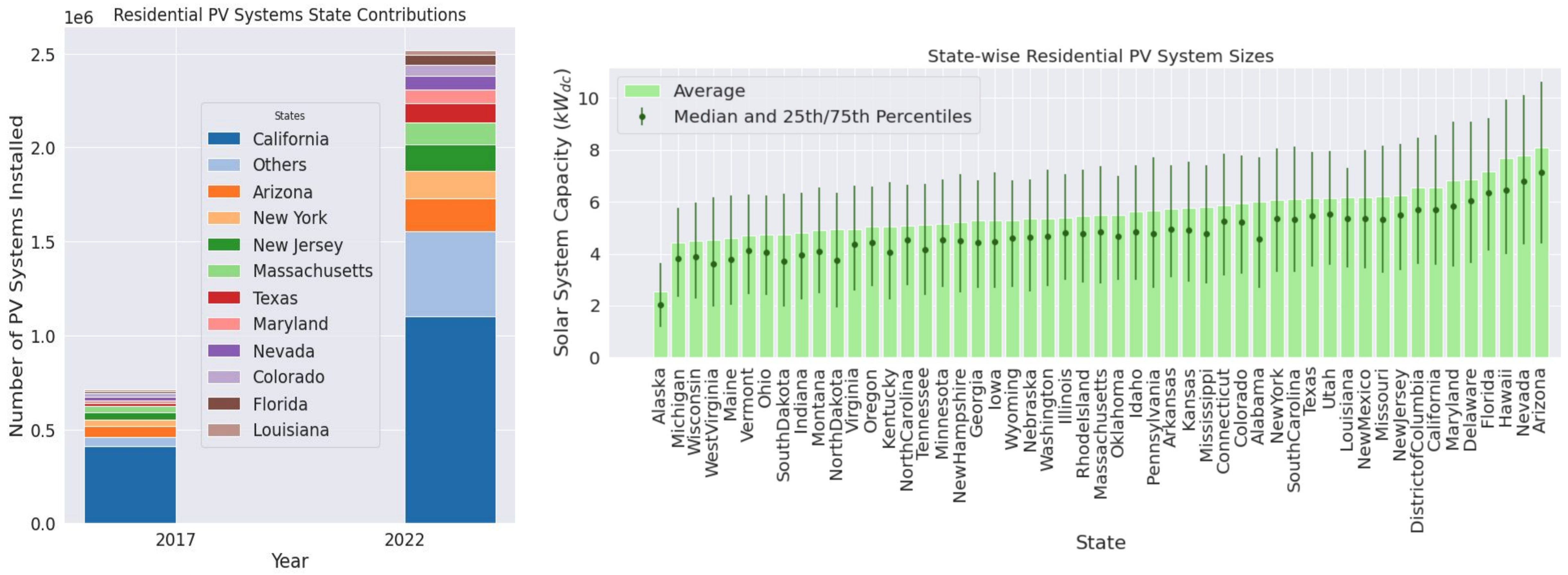

Figure 5. (left) Bar plot comparing Residential PV Systems Count in 2017 and 2022, (right) State-wise Residential PV System Capacities with Average, Median, and 25th/75th Percentiles.

Results

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 in 2022, although 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.

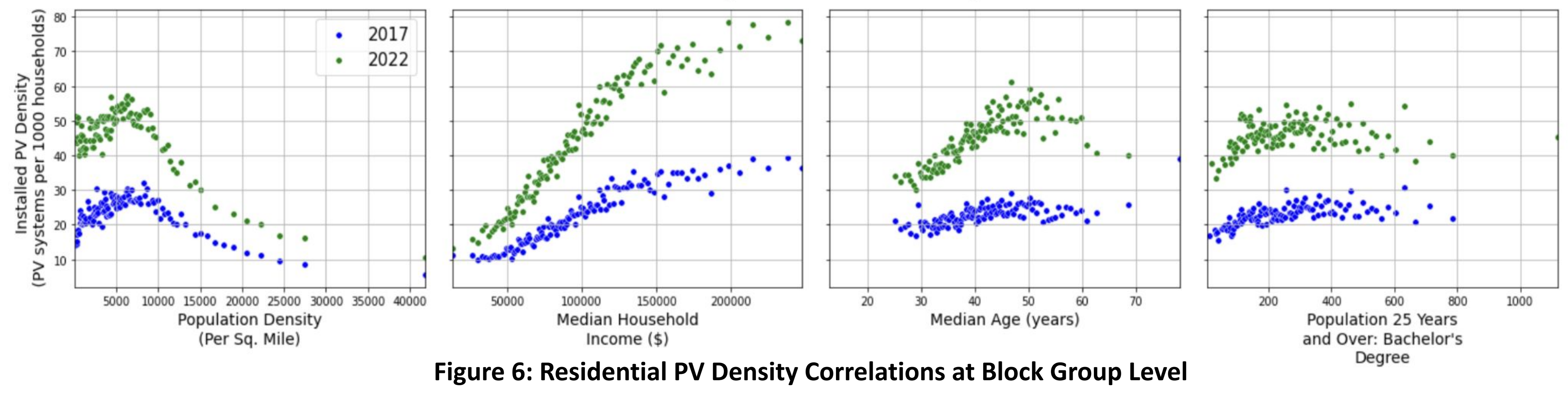

Figure 6: Residential PV Density Correlations at Block Group Level

Conclusion & Future Work

Our AI-powered pipeline maps U.S. residential solar PV systems from 2017–2022, revealing critical equity insights: 50% of block groups now account for 90% of PVs—up from just 20% in 2017, ***20,000+ high-irradiance block groups still have <1% adoption—key zones for community solar, 45% of zero-adoption areas have >40% renters; unlocking these could power 12M homes.*** By translating raw imagery into actionable insights, DeepSolar-3M enables smarter, fairer climate action. We aim to leverage this pipeline to create fast solar PV maps across locations with similar resolution imagery available.

References

[1]. Yu, J., Wang, Z., Majumdar, A., & Rajagopal, R. (2018). DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule*
[2]. Barbose, G. L, Darghouth, N. R, O'Shaughnessy, E., & Forrester, S. (2022). Tracking the Sun. 2022 Edition

Dataset *Reach out!*