Stanford
School of Engineering &
Doerr School of Sustainability
Civil & Environmental Engineering

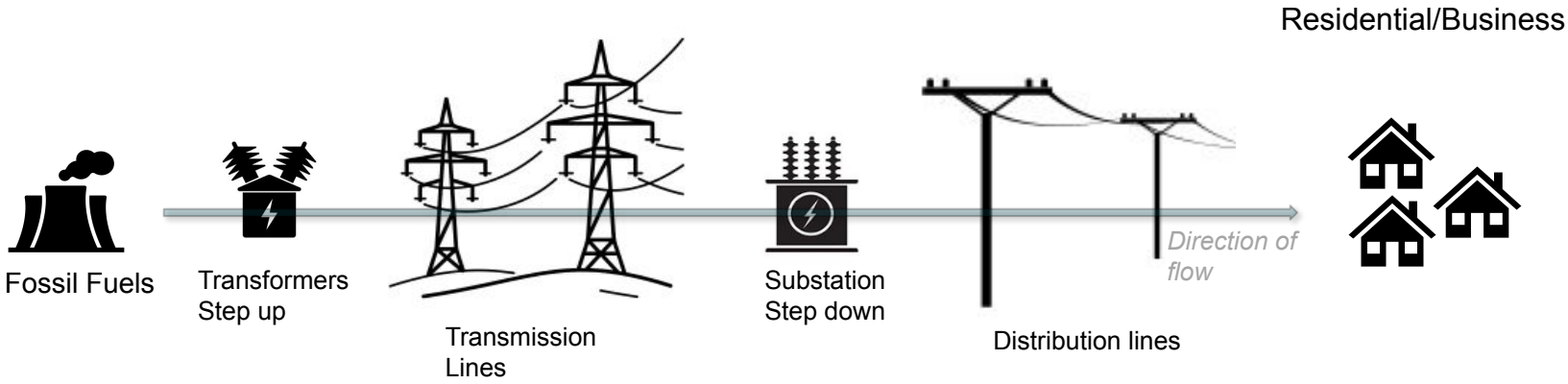
DeepSolar-3M: An AI-Enabled Solar PV Database Documenting 3 Million Systems Across the US

Presented by:

RAJANIE PRABHA

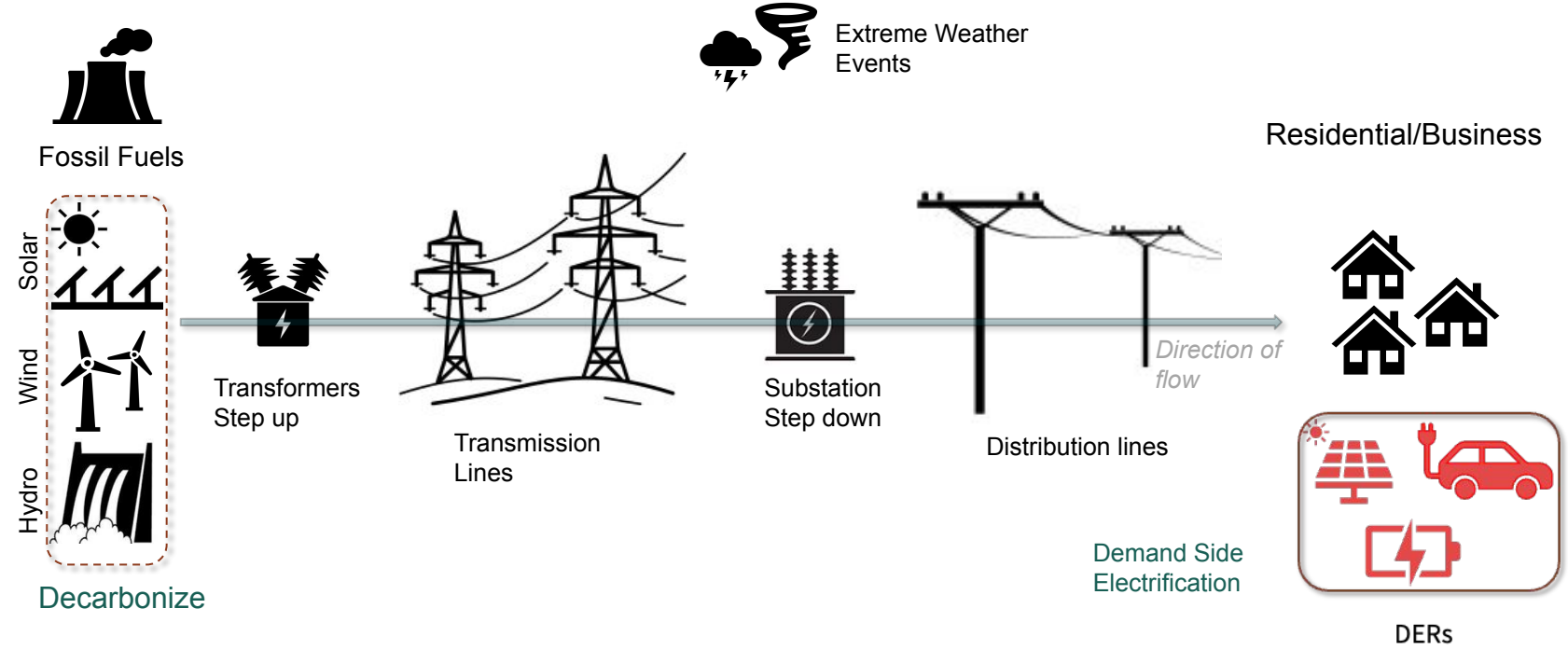
Joint work with Zhecheng Wang, Chad Zanocco, June Flora, Ram Rajagopal

The grid of the past



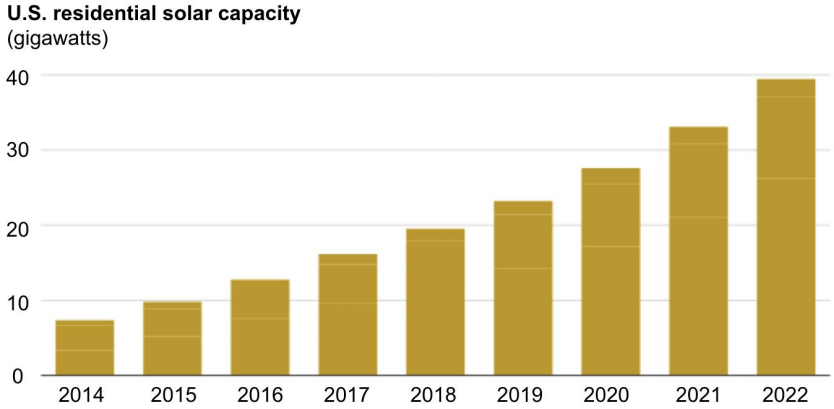
Generation ***Transmission*** ***Distribution*** ***Customer***

The grid of the present



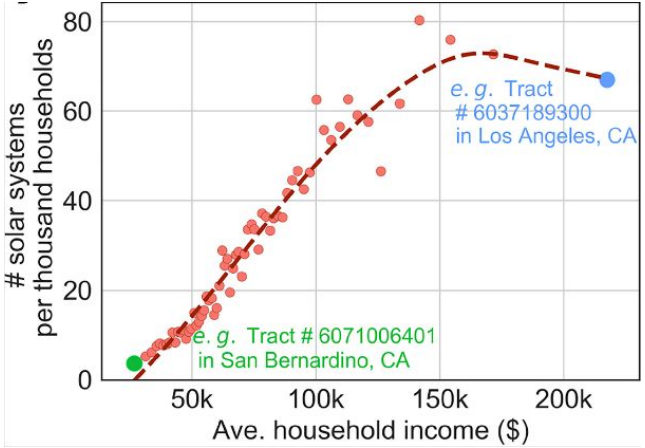
U.S. solar capacity grew by 250% (2017–2022)

Los Angeles, California



Source: Yu 2018, Energy Information Administration (EIA)

However, solar adoption is highly uneven!



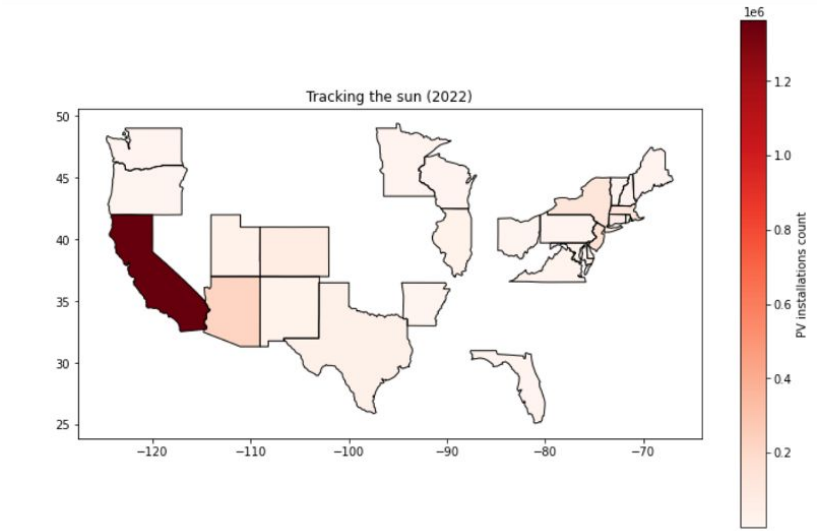
Low- and medium-income households have low deployment densities (2017 installations)

Several policies have been put in place to ensure equitable adoption, however, up-to-date installation data in necessary to ensure that the equity gap is decreasing/the policies are working…..

Source: Yu (2018)

Inadequate Data Acquisition: Conventional Approaches

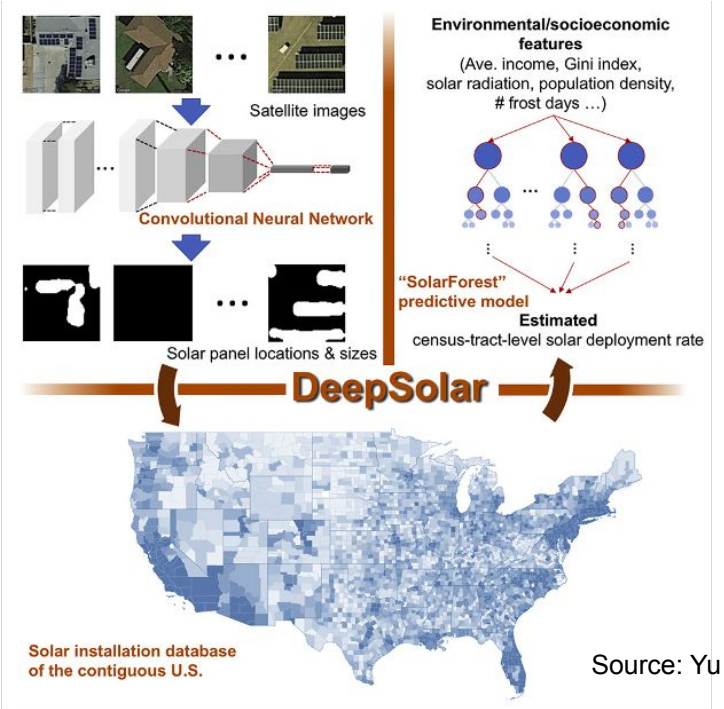
Tracking the Sun (TTS)



- Contributed by agencies, utilities, and other organizations, for systems participating in incentive/net metering programs
- Not scalable, incomplete
- Only covers 31 states, excludes off-grid PV systems

BERKELEY LAB

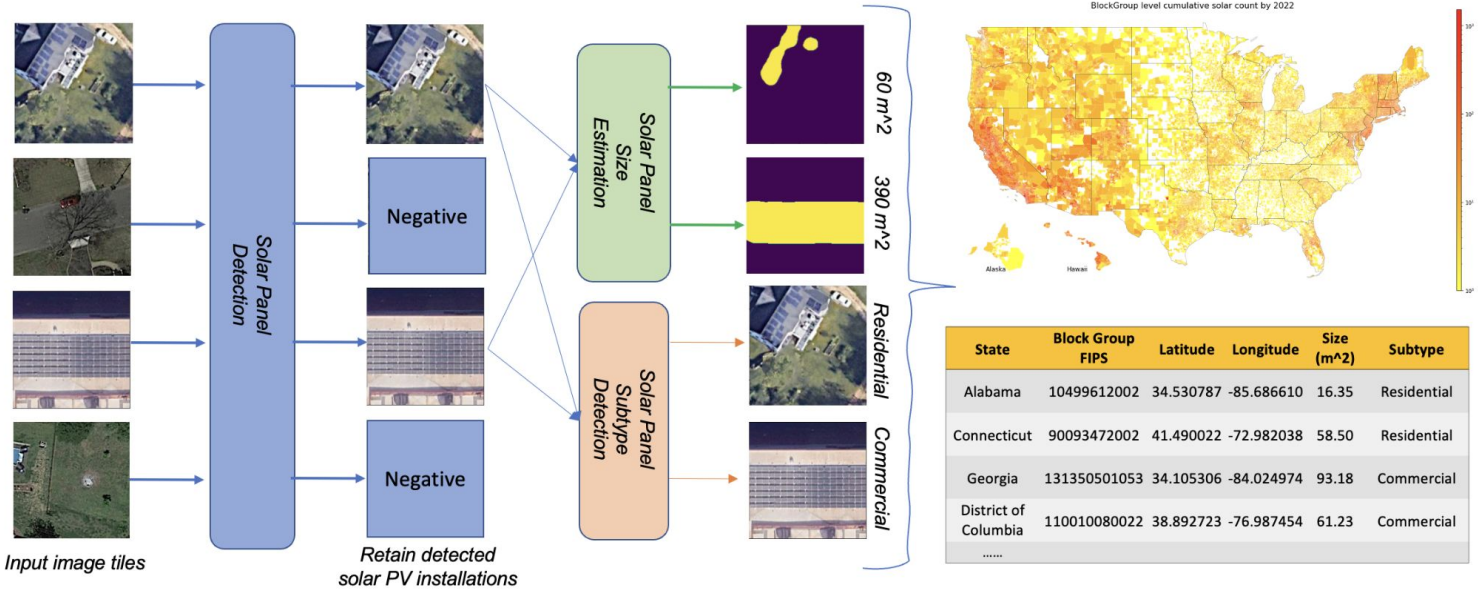
Inadequate Data Acquisition: AI-Based Approaches



Source: Yu (2018)

- Database from 2017
- Pipeline not cost-effective
- Not adaptable to newer Imagery

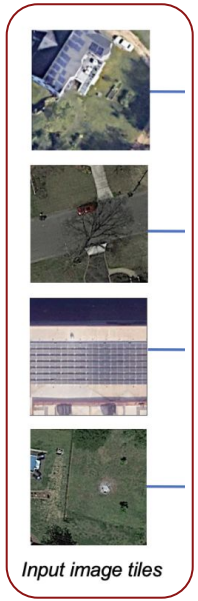
Our Approach

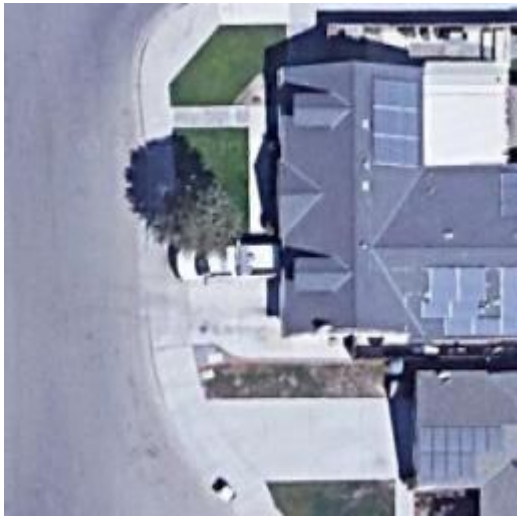


State	Block Group FIPS	Latitude	Longitude	Size (m^2)	Subtype
Alabama	10499612002	34.530787	-85.686610	16.35	Residential
Connecticut	90093472002	41.490022	-72.982038	58.50	Residential
Georgia	131350501053	34.105306	-84.024974	93.18	Commercial
District of Columbia	110010080022	38.892723	-76.987454	61.23	Commercial
.....					

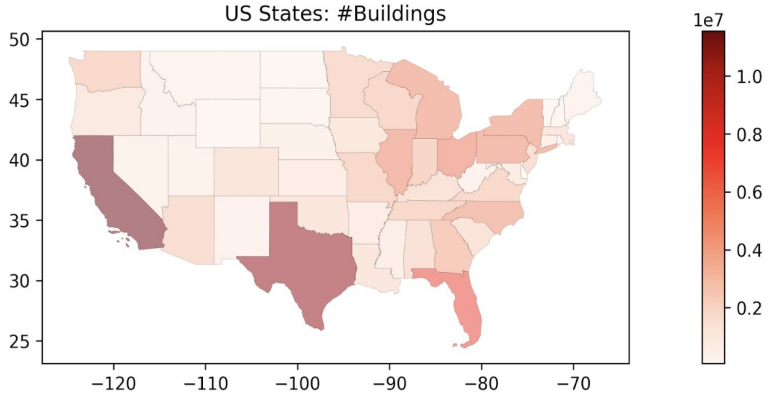
Granular | Scalable | Cost-efficient

Collected ~225 million image tiles



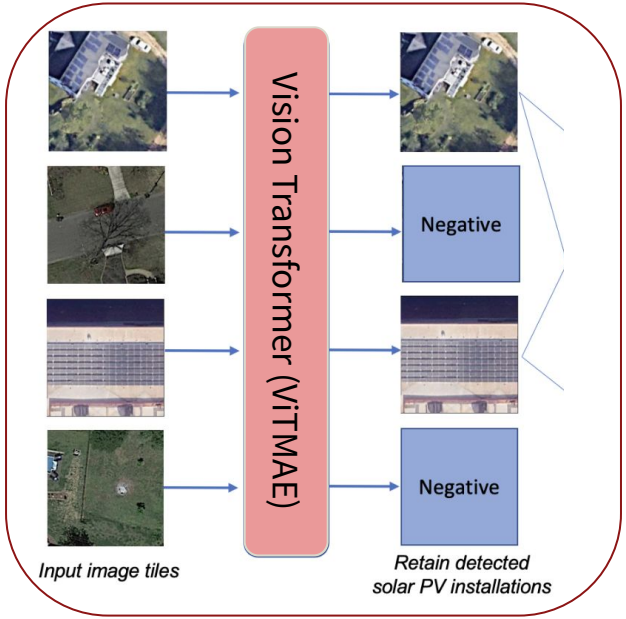


Distribution of Image tiles downloaded across the US



Source: US Building Footprints, Microsoft Maps

The *Detection* Module

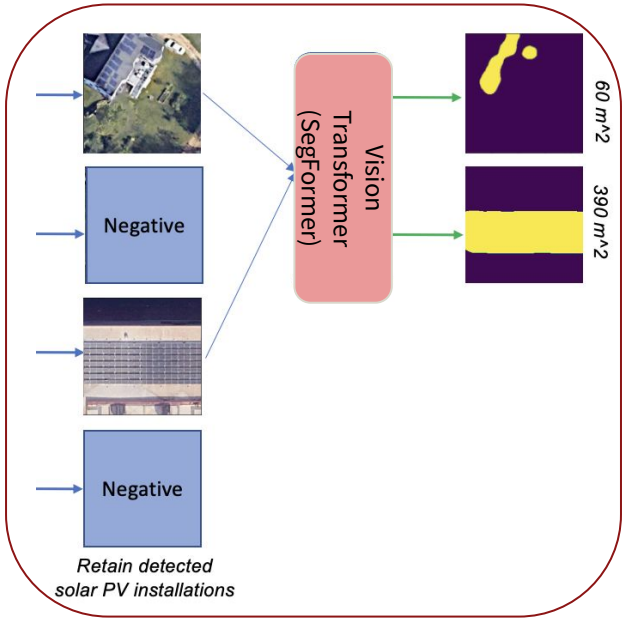


- Fine tuned *ViT* for PV/no-PV binary classification task
- *LoRA* performed the best

Model	# Training Labels	Test set (F1 score)
Deepsolar (CNN)	360k	0.91
Deepsolar-3M (Transformer)	**45k**	**0.92**

Source: He 2021 (VitMAE), Hu 2021 (LoRA)

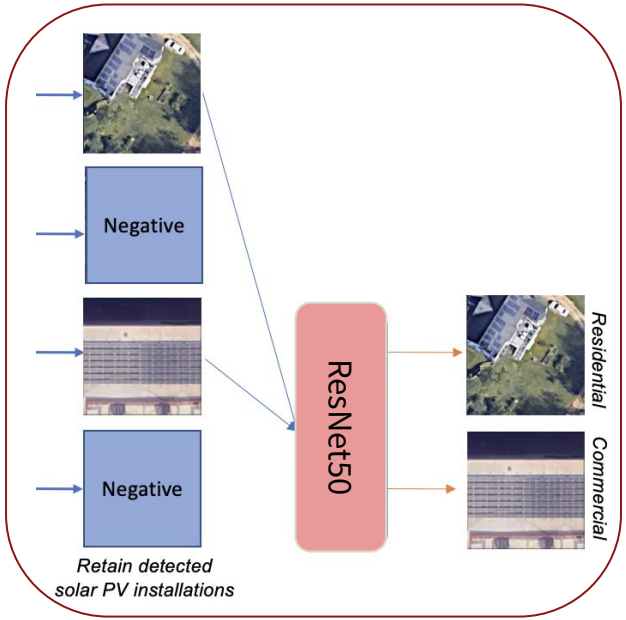
The *Segmentation* Module



- Fine tuned *SegFormer* on 5,607 training labels for semantic segmentation

Model	Test set (size)	Test set (IOU)
Deepsolar-3M (SegFormer)	**600**	**0.92**

The *type-classification* Module



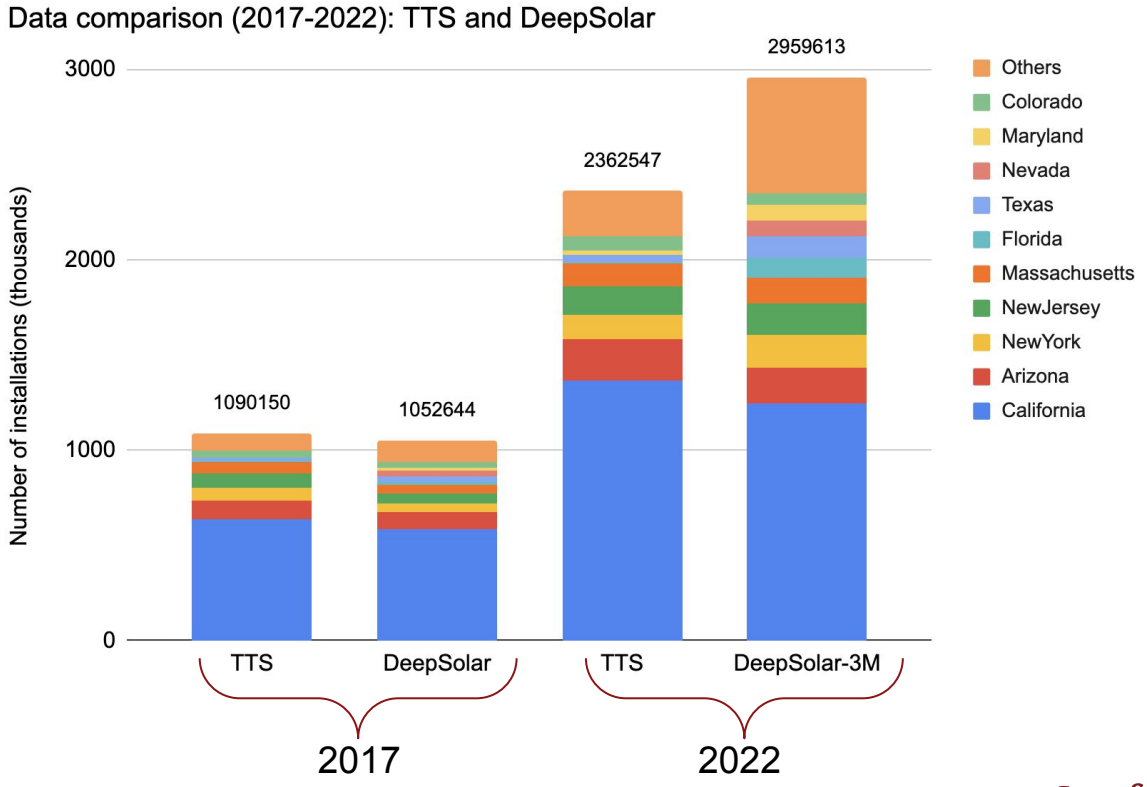
- Fine tuned *Resnet50* on 10,007 training labels for multi-class segmentation: Utility, commercial, residential and solar water heating system

Deepsolar-3M (ResNet)	Test set (size)	Test set (F1)
Residential	**708**	**0.95**
Commercial	**69**	**0.72**
Utility scale	**17**	**0.88**
Solar water heater	**104**	**0.80**

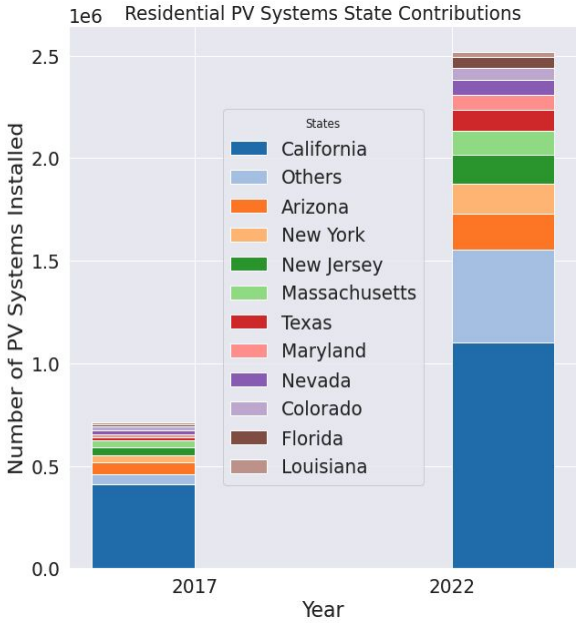
A comprehensive rooftop solar PV map!

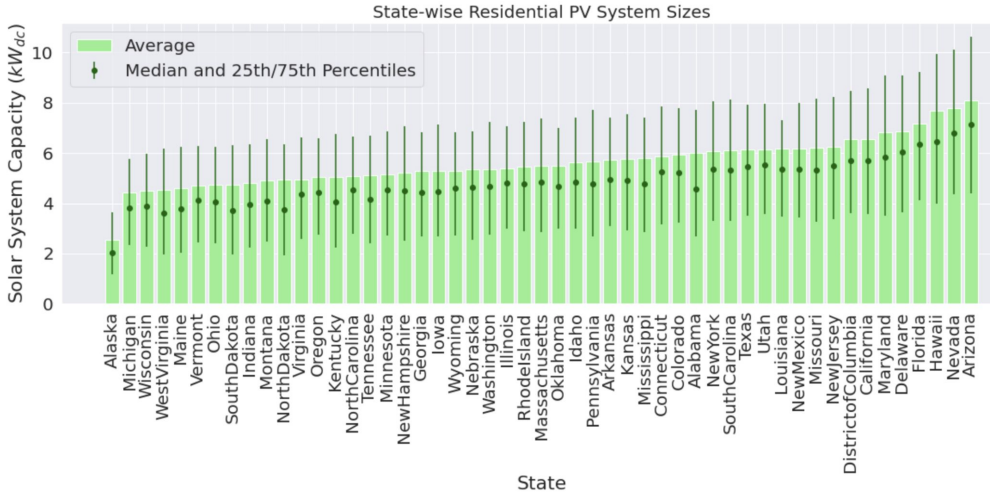
~3 million systems identified

Mapping the overlooked

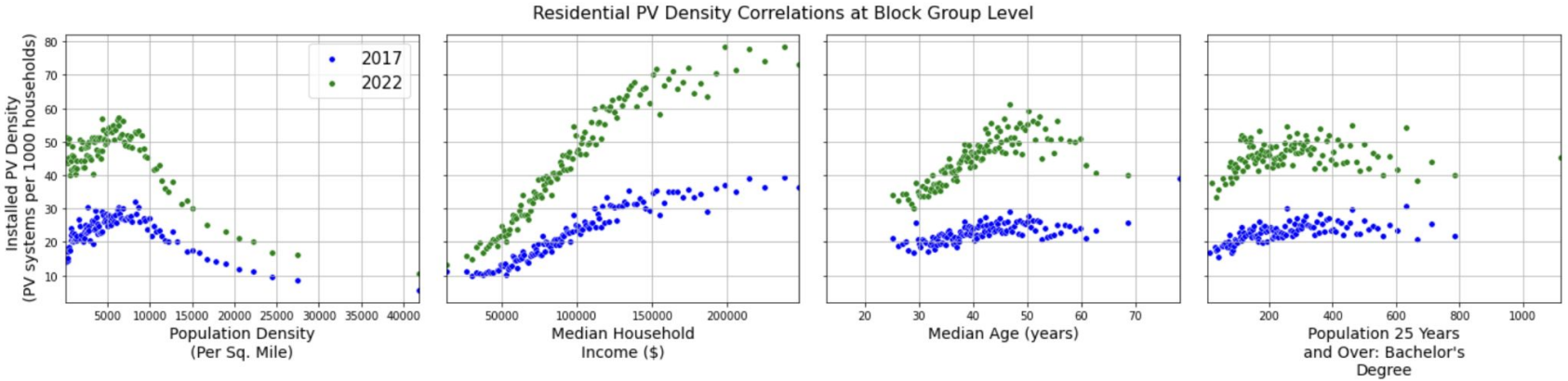


Residential Solar Deployment by State





The equity gap is narrowing..



Why This Dataset Matters

Policymakers
Identify underserved regions, target solar incentives for equity

Utilities & Grid Operators
Forecast and plan distributed generation, grid upgrades

Community & Nonprofits
Advance solar access in marginalized communities, solar energy justice campaigns

Researchers & Data Scientists
Study adoption trends and socioeconomic links, model future growth

Acknowledgements

Stanford ENGINEERING | Civil & Environmental Engineering

SOLAR ENERGY TECHNOLOGIES OFFICE
U.S. Department Of Energy

Stanford
School of Engineering &
Doerr School of Sustainability
Civil & Environmental Engineering

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

TCCML Workshop @ ICLR 2025