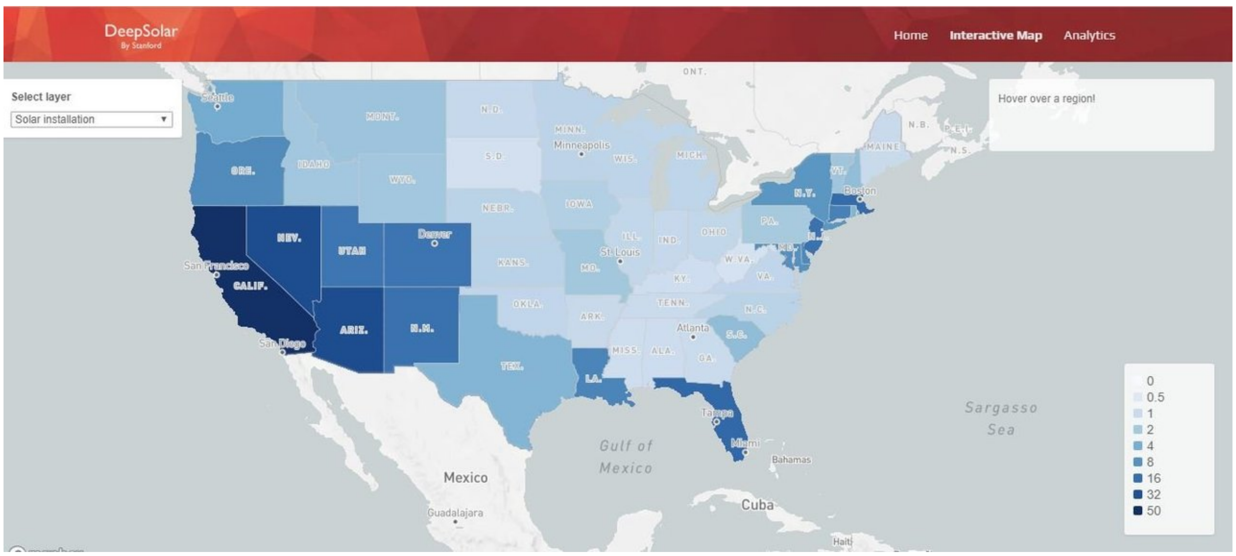
Team locates nearly all US solar panels in a billion images with machine learning

December 19 2018



The interactive map of the United States on the DeepSolar website. Credit: DeepSolar/Stanford University

Knowing which Americans have installed solar panels on their roofs and why they did so would be enormously useful for managing the changing U.S. electricity system and to understanding the barriers to greater use of renewable resources. But until now, all that has been available are essentially estimates.

To get accurate numbers, Stanford University scientists analyzed more than a billion high-resolution satellite images with a machine learning algorithm and identified nearly every solar power installation in the contiguous 48 states. The results are described in a paper published in the Dec. 19 issue of *Joule*. The data are publicly available on the project's website.

The analysis found 1.47 million installations, which is a much higher figure than either of the two widely recognized estimates. The scientists also integrated U.S. Census and other data with their solar catalog to identify factors leading to solar power adoption.

"We can use recent advances in machine learning to know where all these assets are, which has been a huge question, and generate insights about where the grid is going and how we can help get it to a more beneficial place," said Ram Rajagopal, associate professor of civil and environmental engineering, who supervised the project withArun Majumdar, professor of mechanical engineering.

Who goes solar

The group's data could be useful to utilities, regulators, solar panel marketers and others. Knowing how many solar panels are in a neighborhood can help a local electric utility balance supply and demand, the key to reliability. The inventory highlights activators and impediments to solar deployment. For example, the researchers found that household income is very important, but only to a point. Above $150,000 a year, income quickly ceases to play much of a role in people's decisions.

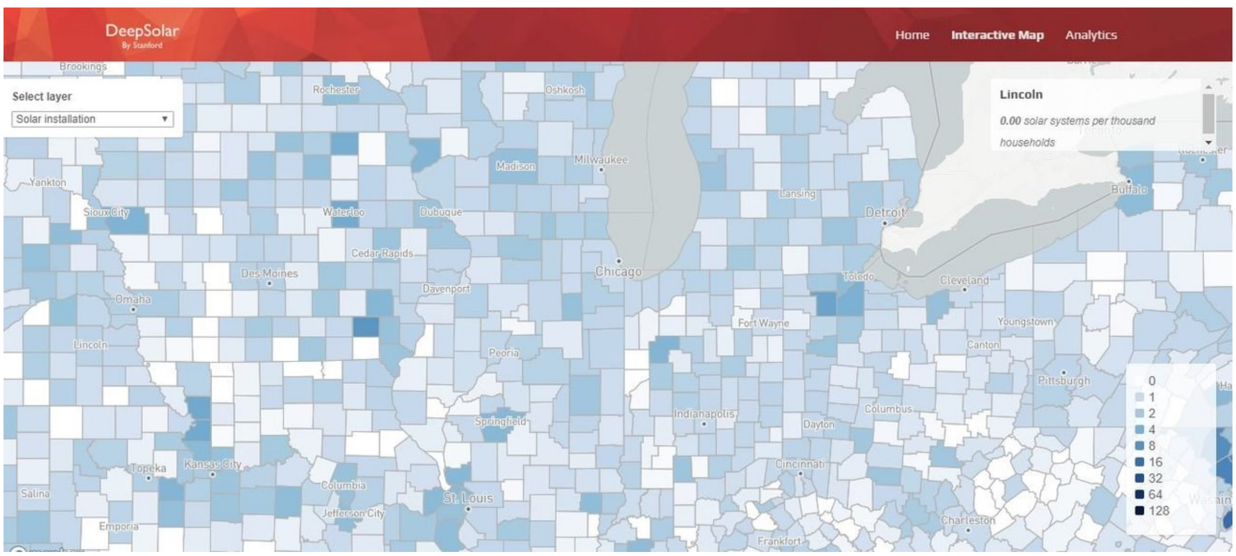
techxplore

This image of the DeepSolar interactive map shows solar panel distribution by county in the San Francisco Bay Area. Credit: DeepSolar/Stanford University

On the other hand, low- and medium-income households do not often install solar systems even when they live in areas where doing so would be profitable in the long term. For example, in areas with a lot of sunshine and relatively high electricity rates, utility bill savings would exceed the monthly cost of the equipment. The impediment for low- and medium-income households is upfront cost, the authors suspect. This finding shows that solar installers could develop new financial models to satisfy unmet demand.

To overlay socioeconomic factors, the team members used publicly available data for U.S. Census tracts. These tracts on average cover about 1,700 households each, about half the size of a ZIP code and about 4 percent of a typical U.S. county. They unearthed other nuggets. For example, once solar penetration reaches a certain level in a neighborhood it takes off, which is not surprising. But if a given neighborhood has a lot of income inequality, that activator often does not switch on. Using geographic data, the team also discovered a significant threshold of how much sunlight a given area needs to trigger adoption.

"We found some insights, but it's just the tip of the iceberg of what we think other researchers, utilities, solar developers and policymakers can further uncover," Majumdar said. "We are making this public so that others find solar deployment patterns, and build economic and behavioral models."



This image of the DeepSolar interactive map shows solar panel distribution by county in the region surrounding Chicago. Credit: DeepSolar/Stanford University

Finding the panels

The team trained the machine learning program, named DeepSolar, to identify solar panels by providing it about 370,000 images, each covering about 100 feet by 100 feet. Each image was labelled as either having or not having a solar panel present. From that, DeepSolar learned to identify features associated with solar panels—for example, color, texture and size.

"We don't actually tell the machine which visual feature is important," said Jiafan Yu, a doctoral candidate in electrical engineering who built the system with Zhecheng Wang, a doctoral candidate in civil and environmental engineering. "All of these need to be learned by the machine."

Eventually, DeepSolar could correctly identify an image as containing solar panels 93 percent of the time and missed about 10 percent of images that did have solar installations. On both scores, DeepSolar is more accurate than previous models, the authors say in the report.

The group then had DeepSolar analyze the billion satellite images to find solar installations—work that would have taken existing technology years to complete. With some novel efficiencies, DeepSolar got the job done in a month.

The resulting database contains not only residential solar installations, but those on the roofs of businesses, as well as many large, utility-owned solar power plants. The scientists, however, had DeepSolar skip the most sparsely populated areas, because it is very likely that buildings in these rural areas either do not have solar panels, or they do but are not attached to the grid. The scientists estimated based on their data that 5 percent of residential and commercial solar installations exist in the areas not covered.

"Advances in machine learning technology have been amazing," Wang said. "But off-the-shelf systems often need to be adapted to the specific project and that requires expertise in the project's topic. Jiafan and I both focus on using the technology to enable renewable energy."

Moving forward, the researchers plan to expand the DeepSolar database to include solar installations in rural areas and in other countries with high-resolution satellite images. They also intend to add features to calculate a solar installation's angle and orientation, which could accurately estimate its power generation. DeepSolar's measure of size is for now only a proxy for potential output.

The group expects to update the U.S. database annually with new satellite images. The information could ultimately feed into efforts to optimize regional U.S. electricity systems, including Rajagopal and Yu's project to help utilities visualize and analyze distributed energy resources.

More information: *Joule*, Yu & Wang et al.: "DeepSolar: A Machine Learning Framework to Efficiently Construct Solar Deployment Database in the United States" www.cell.com/joule/fulltext/S2542-4351(18)30570-1 DOI: 10.1016/j.joule.2018.11.021

Provided by Stanford University