Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States

Timnit Gebru[a,1], Jonathan Krause[a], Yilun Wang[a], Duyun Chen[a], Jia Deng[b], Erez Lieberman Aiden[c,d,e], and Li Fei-Fei[a]

[a]Artificial Intelligence Laboratory, Computer Science Department, Stanford University, Stanford, CA 94305; [b]Vision and Learning Laboratory, Computer Science and Engineering Department, University of Michigan, Ann Arbor, MI 48109; [c]The Center for Genome Architecture, Department of Genetics, Baylor College of Medicine, Houston, TX 77030; [d]Department of Computer Science, Rice University, Houston, TX 77005; and [e]The Center for Genome Architecture, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved October 16, 2017 (received for review January 4, 2017)

The United States spends more than $250 million each year on the American Community Survey (ACS), a labor-intensive door-to-door study that measures statistics relating to race, gender, education, occupation, unemployment, and other demographic factors. Although a comprehensive source of data, the lag between demographic changes and their appearance in the ACS can exceed several years. As digital imagery becomes ubiquitous and machine vision techniques improve, automated data analysis may become an increasingly practical supplement to the ACS. Here, we present a method that estimates socioeconomic characteristics of regions spanning 200 US cities by using 50 million images of street scenes gathered with Google Street View cars. Using deep learning-based computer vision techniques, we determined the make, model, and year of all motor vehicles encountered in particular neighborhoods. Data from this census of motor vehicles, which enumerated 22 million automobiles in total (8% of all automobiles in the United States), were used to accurately estimate income, race, education, and voting patterns at the zip code and precinct level. (The average US precinct contains ∼1,000 people.) The resulting associations are surprisingly simple and powerful. For instance, if the number of sedans encountered during a drive through a city is higher than the number of pickup trucks, the city is likely to vote for a Democrat during the next presidential election (88% chance); otherwise, it is likely to vote Republican (82%). Our results suggest that automated systems for monitoring demographics may effectively complement labor-intensive approaches, with the potential to measure demographics with fine spatial resolution, in close to real time.

computer vision | deep learning | social analysis | demography

For thousands of years, rulers and policymakers have surveyed national populations to collect demographic statistics. In the United States, the most detailed such study is the American Community Survey (ACS), which is performed by the US Census Bureau at a cost of $250 million per year (1). Each year, ACS reports demographic results for all cities and counties with a population of 65,000 or more (2). However, due to the labor-intensive data-gathering process, smaller regions are interrogated less frequently, and data for geographical areas with less than 65,000 inhabitants are typically presented with a lag of ~ 2.5 y. Although the ACS represents a vast improvement over the earlier, decennial census (3), this lag can nonetheless impede effective policymaking. Thus, the development of complementary approaches would be desirable.

In recent years, computational methods have emerged as a promising tool for tackling difficult problems in social science. For instance, Antenucci et al. (4) have predicted unemployment rates from Twitter; Michel et al. (5) have analyzed culture using large quantities of text from books; and Blumenstock et al. (6) used mobile phone metadata to predict poverty rates in Rwanda. These results suggest that socioeconomic studies, too, might be facilitated by computational methods, with the ultimate potential of analyzing demographic trends in great detail, in real time, and at a fraction of the cost.

Recently, Naik et al. (7) used publicly available imagery to quantify people's subjective perceptions of a neighborhood's physical appearance. They then showed that changes in these perceptions correlate with changes in socioeconomic variables (8). Our work explores a related theme: whether socioeconomic statistics can be inferred from objective characteristics of images from a neighborhood.

Here, we show that it is possible to determine socioeconomic statistics and political preferences in the US population by combining publicly available data with machine-learning methods. Our procedure, designed to build upon and complement the ACS, uses labor-intensive survey data for a handful of cities to train a model that can create nationwide demographic estimates. This approach allows for estimation of demographic variables with high spatial resolution and reduced lag time.

Specifically, we analyze 50 million images taken by Google Street View cars as they drove through 200 cities, neighborhood-by-neighborhood and street-by-street. In Google Street View images, only the exteriors of houses, landscaping, and vehicles on the street can be observed. Of these objects, vehicles are among the most personalized expressions of American culture: Over 90% of American households own a motor vehicle (9), and their choice of automobile is influenced by disparate demographic factors including household needs, personal preferences, and economic wherewithal (10). (Note that, in principle, other factors such as spacing between houses, number of stories, and extent of shrubbery could also be integrated into such models.) Such street scenes are a natural data type to explore: They already cover

Significance

We show that socioeconomic attributes such as income, race, education, and voting patterns can be inferred from cars detected in Google Street View images using deep learning. Our model works by discovering associations between cars and people. For example, if the number of sedans in a city is higher than the number of pickup trucks, that city is likely to vote for a Democrat in the next presidential election (88% chance); if not, then the city is likely to vote for a Republican (82% chance).

Author contributions: T.G., J.K., J.D., E.L.A., and L.F.-F. designed research; T.G., J.K., Y.W., D.C., J.D., E.L.A., and L.F.-F. performed research; T.G. and J.K. contributed new reagents/analytic tools; T.G., J.K., Y.W., D.C., J.D., E.L.A., and L.F.-F. analyzed data; and T.G., J.K., E.L.A., and L.F.-F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: tgebru@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1700035114/-/DCSupplemental.