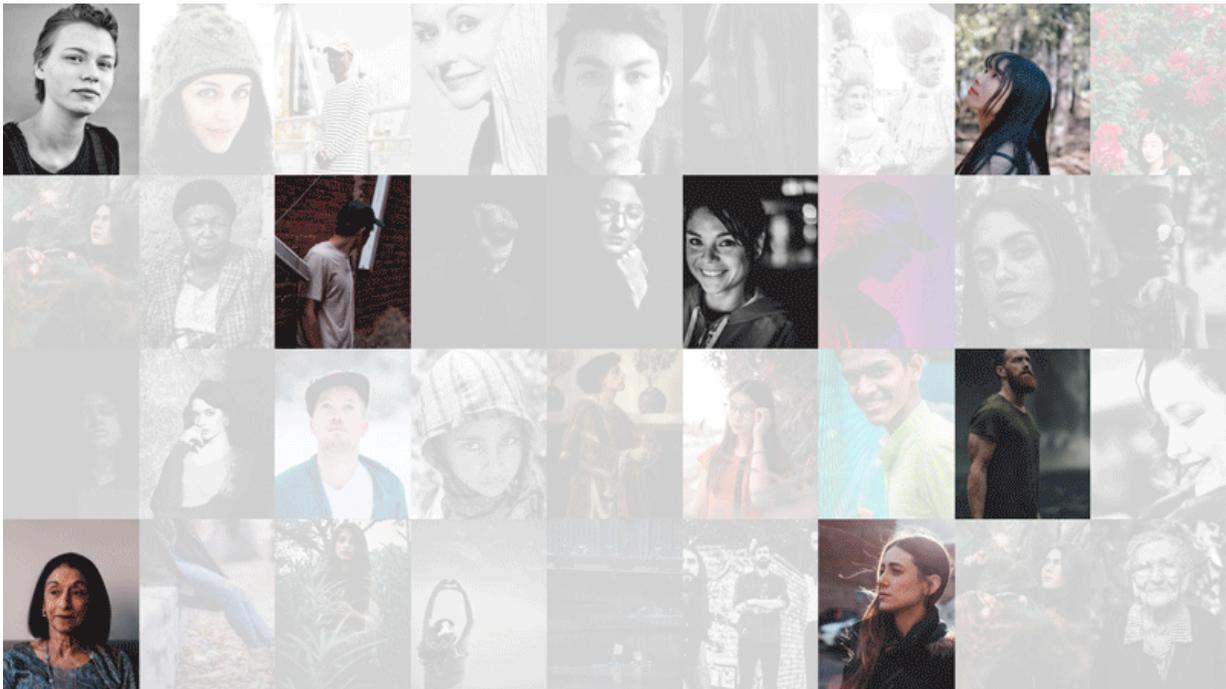


Researchers devise approach to reduce biases in computer vision data sets

February 13 2020, by Molly Sharlach



Addressing problems of bias in artificial intelligence, computer scientists from Princeton and Stanford University have proposed improvements to ImageNet, a database of more than 14 million images. The researchers developed a tool that allows users to specify and retrieve image sets of people that are balanced by age, gender expression or skin color. The above animation is a conceptual representation of the tool. Credit: Ryan Rizzuto

Addressing problems of bias in artificial intelligence, computer scientists

from Princeton and Stanford University have developed methods to obtain fairer data sets containing images of people. The researchers propose improvements to ImageNet, a database of more than 14 million images that has played a key role in advancing computer vision over the past decade.

ImageNet, which includes images of objects and landscapes as well as people, serves as a source of training data for researchers creating machine learning algorithms that classify images or recognize elements within them. ImageNet's unprecedented scale necessitated automated image collection and crowdsourced image annotation. While the database's person categories have rarely been used by the research community, the ImageNet team has been working to address biases and other concerns about images featuring people that are unintended consequences of ImageNet's construction.

"Computer vision now works really well, which means it's being deployed all over the place in all kinds of contexts," said co-author Olga Russakovsky, an assistant professor of computer science at Princeton. "This means that now is the time for talking about what kind of impact it's having on the world and thinking about these kinds of fairness issues."

In a new paper, the ImageNet team systematically identified non-visual concepts and offensive categories, such as racial and sexual characterizations, among ImageNet's person categories and proposed removing them from the database. The researchers also designed a tool that allows users to specify and retrieve image sets of people that are balanced by age, [gender expression](#) or skin color—with the goal of facilitating algorithms that more fairly classify people's faces and activities in images. The researchers presented their work on Jan. 30 at the Association for Computing Machinery's Conference on Fairness, Accountability and Transparency in Barcelona, Spain.

"There is very much a need for researchers and labs with core technical expertise in this to engage in these kinds of conversations," said Russakovsky. "Given the reality that we need to collect the data at scale, given the reality that it's going to be done with crowdsourcing because that's the most efficient and well-established pipeline, how do we do that in a way that's fairer—that doesn't fall into these kinds of prior pitfalls? The core message of this paper is around constructive solutions."

A group of [computer scientists](#) at Princeton and Stanford launched ImageNet in 2009 as a resource for academic researchers and educators. Leading the effort was Princeton alumna and faculty member Fei-Fei Li, now a professor of computer science at Stanford. To encourage researchers to build better computer vision algorithms using ImageNet, the team also created the ImageNet Large Scale Visual Recognition Challenge. The challenge focused largely on object recognition using 1,000 image categories, only three of which featured people.

Some of the fairness issues in ImageNet stem from the pipeline used to build the database. Its image categories came from WordNet, an older database of English words used for natural language processing research. ImageNet's creators adopted the nouns in WordNet—some of which, although they are clearly defined verbal terms, do not translate well to a visual vocabulary. For example, terms that describe a person's religion or geographic origin might retrieve only the most distinctive image search results, potentially leading to algorithms that perpetuate stereotypes.

A recent art project called ImageNet Roulette brought increased attention to these concerns. The project, released in September 2019 as part of an art exhibition on image recognition systems, used images of people from ImageNet to train an artificial intelligence model that classified people in words based on a submitted image. Users could upload an image of themselves and retrieve a label based on this model. Many of the classifications were offensive or simply off-base.

The central innovation that allowed ImageNet's creators to amass such a large database of labeled images was the use of crowdsourcing—specifically, the Amazon Mechanical Turk (MTurk) platform, through which workers were paid to verify candidate images. This approach, while transformative, was imperfect, leading to some biases and inappropriate categorizations.

"When you ask people to verify images by selecting the correct ones from a large set of candidates, people feel pressured to select some images and those images tend to be the ones with distinctive or stereotypical features," said lead author Kaiyu Yang, a graduate student in computer science.

In the study, Yang and colleagues first filtered out potentially offensive or sensitive person categories from ImageNet. They defined offensive categories as those containing profanity or racial or gender slurs; sensitive categories included, for example, the classification of people based on sexual orientation or religion. To annotate the categories, they recruited 12 graduate students from diverse backgrounds, instructing them to err on the side of labeling a category as sensitive if they were unsure. This eliminated 1,593 categories—about 54% of the 2,932 person categories in ImageNet.

The researchers then turned to MTurk workers to rate the "imageability" of the remaining safe categories on a scale of one to five. Keeping categories with an imageability rating of four or higher resulted in only 158 categories classified as both safe and imageable. Even this highly filtered set of categories contained more than 133,000 images—a wealth of examples for training computer vision algorithms.

Within these 158 categories, the researchers studied the demographic representation of people in the images in order to assess the level of bias in ImageNet and devise an approach to create fairer data sets.

ImageNet's content comes from image search engines such as Flickr, and search engines in general have been shown to produce results that overrepresent males, light-skinned people, and adults between the ages of 18 and 40.

"People have found that the distributions of demographics in image search results are highly biased, and this is why the distribution in ImageNet is also biased," said Yang. "In this paper we tried to understand how biased it is, and also to propose a method to balance the distribution."

Of the attributes protected under U.S. anti-discrimination laws, the researchers considered the three attributes that are imageable: skin color, gender expression and age. MTurk workers were asked to annotate each attribute of each person in an image. They classified [skin color](#) as light, medium or dark; and age as child (under 18), adult 18–40, adult 40–65 or adult over 65. Gender classifications included male, female and unsure—a way to include people with diverse gender expressions, as well as annotate images in which gender could not be perceived from visual clues (such as many images of babies or scuba divers).

An analysis of the annotations showed that, similar to search results, ImageNet's content reflects considerable bias. People annotated as dark-skinned, females, and adults over 40 were underrepresented across most categories.

Although the annotation process included quality controls and required annotators to reach consensus, out of concern for the potential harm of mis-annotations, the researchers opted not to release demographic annotations for individual images. Instead, they designed a web-interface tool that allows users to obtain a set of images that are demographically balanced in a way the user specifies. For example, the full collection of images in the category "programmer" may include about 90% males and

10% females, while in the United States about 20% of computer programmers are female. A researcher could use the new tool to retrieve a set of programmer images representing 80% males and 20% females—or an even split, depending on the researcher's purpose.

"We do not want to say what is the correct way to balance the demographics, because it's not a very straightforward issue," said Yang. "The distribution could be different in different parts of the world—the distribution of skin colors in the U.S. is different than in countries in Asia, for example. So we leave that question to our user, and we just provide a tool to retrieve a balanced subset of the images."

The ImageNet team is currently working on technical updates to its hardware and database, in addition to implementing the filtering of the person categories and the rebalancing tool developed in this research. ImageNet will soon be re-released with these updates, and with a call for feedback from the computer vision research community.

More information: Towards fairer datasets: filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency doi.org/10.1145/3351095.3375709

Provided by Princeton University

Citation: Researchers devise approach to reduce biases in computer vision data sets (2020, February 13) retrieved 23 April 2024 from <https://techxplore.com/news/2020-02-approach-biases-vision.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.