Using deep learning to predict users' superficial judgments of human faces

April 22 2022, by Ingrid Fadelli

Credit: Peterson et al.

Many psychology studies have confirmed the biased nature of human judgments and decision-making. When interacting with a new person, for instance, humans often make a series of automatic and superficial judgments based solely on their appearance, facial features, ethnicity, body-type, and body language.

Researchers at Researchers at Princeton University, Stevens Institute of Technology, and the Booth School of Business of the University of Chicago have recently tried to predict some of the automatic inferences that humans make about others based solely on their face, using deep neural networks. Their paper, published in *PNAS*, introduces a machine learning model that can predict the arbitrary judgments users will make about specific pictures of faces with remarkable accuracy.

"As psychologists, we are interested in how people perceive and judge faces, especially when there are important consequences, such as hiring and sentencing decisions involved," Joshua Peterson, one of the researchers who carried out the study, told TechXplore "However, most work up to now was limited to studying artificial 3D face renderings or small sets of photographs."

In recent years, computer scientists have developed a wide range of advanced machine learning models that can analyze and classify large amounts of data, predict specific events with good accuracy, and generate images, audio recordings or texts. While reviewing previous literature focusing on human face judgments, however, Peterson and his colleagues noticed that very few studies explored this topic using state-of-the-art machine learning tools.

"The main objective of our recent study was to produce a scientific model of people's impressions of faces that generalized to as many possible faces and attributes (e.g., trustworthiness) as possible," Peterson said. "We also wanted the model to double as a tool for generating and

manipulating face stimuli in psychology, and also represent a new standard for studying arbitrary attribute inferences for faces."
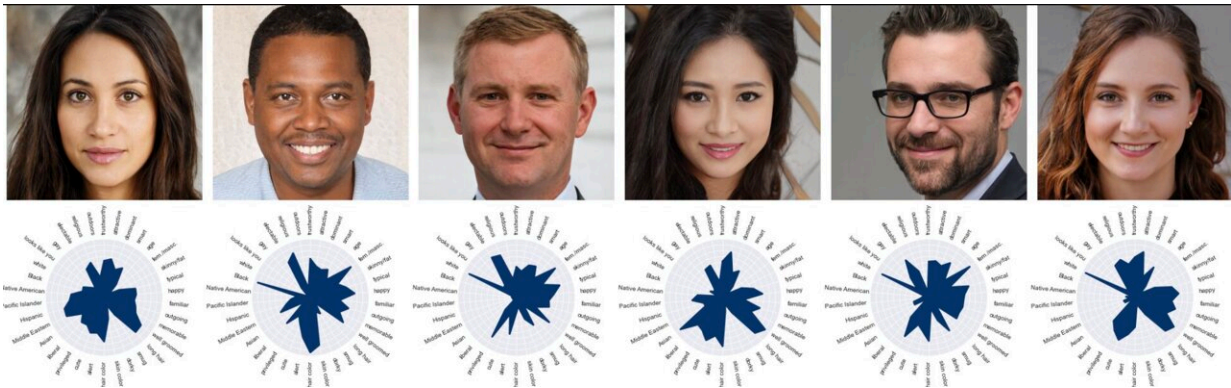
Multiple examples of faces from the dataset and average ratings for each attribute Credit: Peterson et al.

In the past, deep neural network-based models were primarily used to automatically detect facial expressions, basic emotions, or the presence of specific accessories (e.g., glasses, sunglasses, earrings, etc.). Peterson and his colleagues, on the other hand, wished to use deep neural networks to model personality-related attributes that humans might typically infer from faces, such as trustworthiness.

Face-related inferences are biased and arbitrary, which means that they can be completely different based on who is making them. Therefore, to successfully model them, the researchers first had to compile a large dataset containing both images of faces and the judgments that many different humans made about them.

"Even though our dataset is the largest of its kind in psychology, containing over 1 million judgments, it's still not big enough to train a

neural network model from scratch," Peterson said. "Instead, we assume that existing models have already adequately learned the general structure of faces from a larger unlabeled dataset, and then we can aim our own behavioral data directly at the remaining problem of relating that structure to psychological inferences."

Instead of learning the weights of an entirely new neural network, which is what deep learning models for the analysis of faces are typically programmed to do, the model developed by Peterson and his colleagues specifically learns weights that associate judgments of trustworthiness to facial features that were already uncovered by another existing model. This ultimately allowed the researchers to interpret their deep neural network's opaque internal states from a psychological standpoint.

"As the base network we chose is generative, we are also able to manipulate faces along these interpretable dimensions, such that they will be judged as more or less e.g., trustworthy, which we verified in separate behavioral experiments," Peterson said.

The recent work by this team of researchers has led to the creation of what may be the most comprehensive and detailed dataset containing face-related biases and stereotypes compiled to date. In the future, this dataset and the deep [neural network](#) presented in their paper could be used to investigate these biases further, particularly in contexts such as professional recruiting and criminal law cases. In addition, they could guide the development of more effective strategies to reduce the impact of such biases.

"The current dataset that powers our model consists of judgments from a mostly White, North American population," Peterson said. "One important extension of the work will be to ask how the biases we are studying differ across much more diverse populations. Anyone who wants to help us do this can participate in our study by judging faces at

https://demo.onemillionimpressions.com/v2/consent/."

   More information: Joshua C. Peterson et al, Deep models of superficial face judgments, *Proceedings of the National Academy of Sciences* (2022). DOI: 10.1073/pnas.2115228119

© 2022 Science X Network