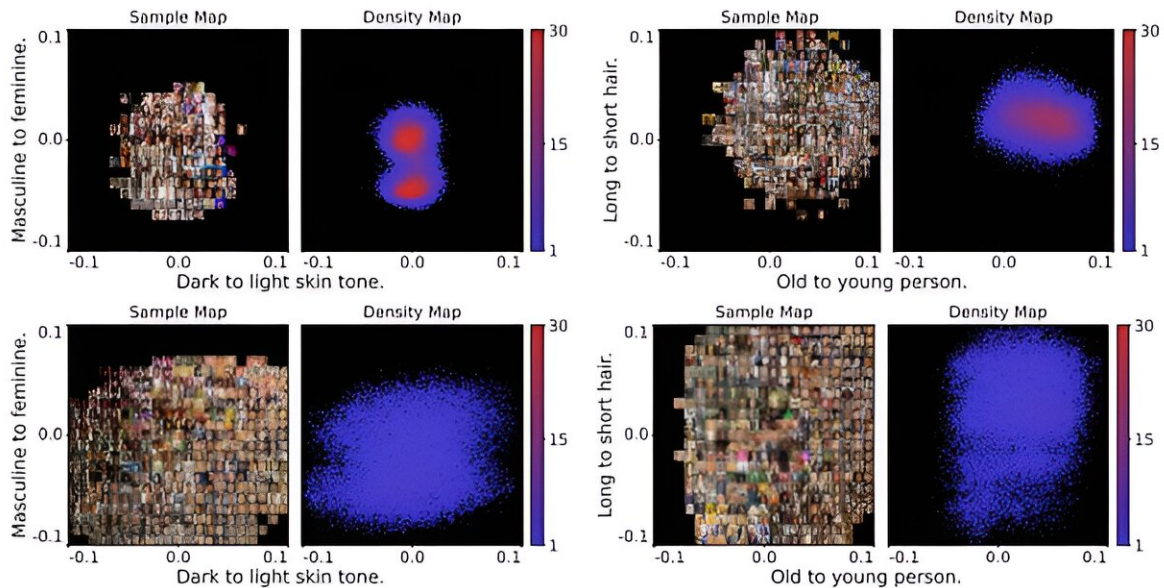


Diversifying data to beat bias

February 23 2024, by Caitlin Dawson and Anna Hsu



Sample and density maps of synthetic datasets projected onto CLIP latent planes, computed via measure functions described in Section 5.1. Top: Random sampling (Rand50) creates high-density clusters and have sparse representation of extreme measures. Bottom: QDGS (QD50) can create more uniform distributions of synthetic data across measures. Credit: *arXiv* (2023). DOI: 10.48550/arxiv.2312.14369

AI holds the potential to revolutionize health care, but it also brings with it a significant challenge: bias. For instance, a dermatologist might use an AI-driven system to help identify suspicious moles. But what if the machine learning model was trained primarily on image data from

lighter skin tones, and misses a common form of skin cancer on a darker-skinned patient?

This is a real-world problem. In 2021, [researchers found that](#) free image databases that could be used to train AI systems to diagnose [skin cancer](#) contain very few images of people with darker skin. It turns out, AI is only as good as its data, and biased data can lead to serious outcomes, including unnecessary surgery and even missing treatable cancers.

In a new paper presented at the AAAI Conference on Artificial Intelligence, USC computer science researchers propose a novel approach to mitigate bias in [machine learning model](#) training, specifically in image generation.

The researchers used a family of algorithms, called "quality-diversity algorithms" or QD algorithms, to create diverse synthetic datasets that can strategically "plug the gaps" in real-world training data.

The paper, titled "[Quality-Diversity Generative Sampling for Learning with Synthetic Data](#)," appears on the pre-print server *arXiv* and was lead-authored by Allen Chang, a senior double majoring in computer science and applied math.

"I think it is our responsibility as computer scientists to better protect all communities, including minority or less frequent groups, in the systems we design," said Chang. "We hope that quality-diversity optimization can help to generate fair synthetic data for broad impacts in [medical applications](#) and other types of AI systems."

Increasing fairness

While generative AI models have been used to create synthetic data in the past, "there's a danger of producing biased data, which can further

bias downstream models, creating a vicious cycle," said Chang.

Quality diversity algorithms, on the other hand, are typically used to generate diverse solutions to a problem, for instance, helping robots explore unknown environments, or generating game levels in a video game. In this case, the algorithms were put to work in a new way: to solve the problem of creating diverse synthetic datasets.

Using this method, the team was able to generate a diverse dataset of around 50,000 images in 17 hours, around 20 times more efficiently than traditional methods of "rejection sampling," said Chang. The team tested the dataset on up to four measures of diversity—[skin tone](#), gender presentation, age, and hair length.

"We found that training data produced with our method has the potential to increase fairness in the machine learning model, increasing accuracy on faces with darker skin tones, while maintaining accuracy from training on additional data," said Chang.

"This is a promising direction for augmenting models with bias-aware sampling, which we hope can help AI systems perform accurately for all users."

Notably, the method increases the representation of intersectional groups—a term for groups with multiple identities—in the data. For instance, people who have both dark skin tones and wear eyeglasses, which would be especially limited traits in traditional real-world datasets.

"While there has been previous work on leveraging QD algorithms to generate diverse content, we show for the first time that generative models can use QD to repair biased classifiers," said Nikolaidis.

"They do this by iteratively generating and rebalancing content across

user-specified features, using the newly balanced content to improve classifier fairness. This work is a first step in the direction of enabling biased models to 'self-repair' by iteratively generating and retraining on synthetic data."

More information: Allen Chang et al, Quality-Diversity Generative Sampling for Learning with Synthetic Data, *arXiv* (2023). [DOI: 10.48550/arxiv.2312.14369](https://doi.org/10.48550/arxiv.2312.14369)

Provided by University of Southern California

Citation: Diversifying data to beat bias (2024, February 23) retrieved 30 April 2024 from <https://techxplore.com/news/2024-02-diversifying-bias.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.