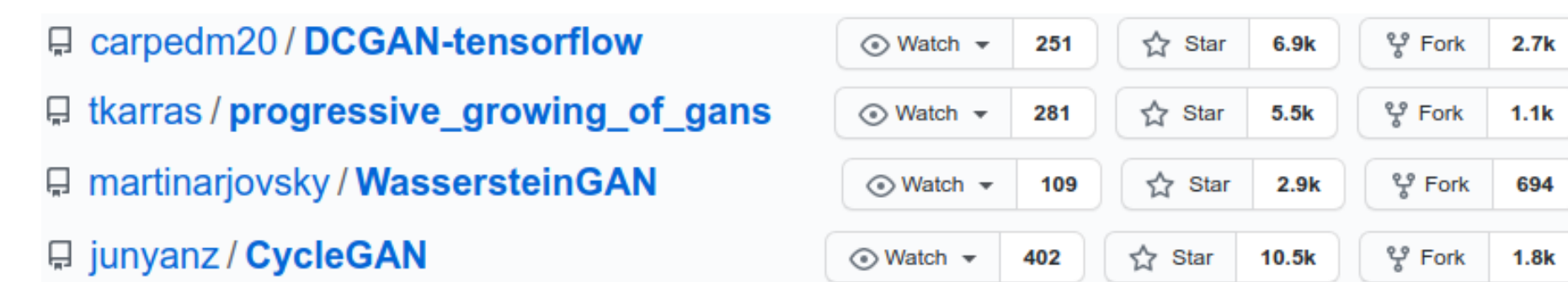


Imperfect ImaGANation: Implications of GANs Exacerbating Biases on Facial Data

Niharika Jain* (🐦), Alberto Olmo* (🐦), Sailik Sengupta (🐦), Lydia Manikonda (🐦), Subbarao Kambhampati (🐦)

Motivation



- Wide adoption of GANs as a seemingly trustworthy data augmentation technique.
- Practitioners possibly unaware of Mode Collapse causing exacerbation of biases.

The Mode Collapse Problem

The diversity of the generated distribution is much lower than that of the training set due to the non-infinite capacity of the generator nor discriminator.

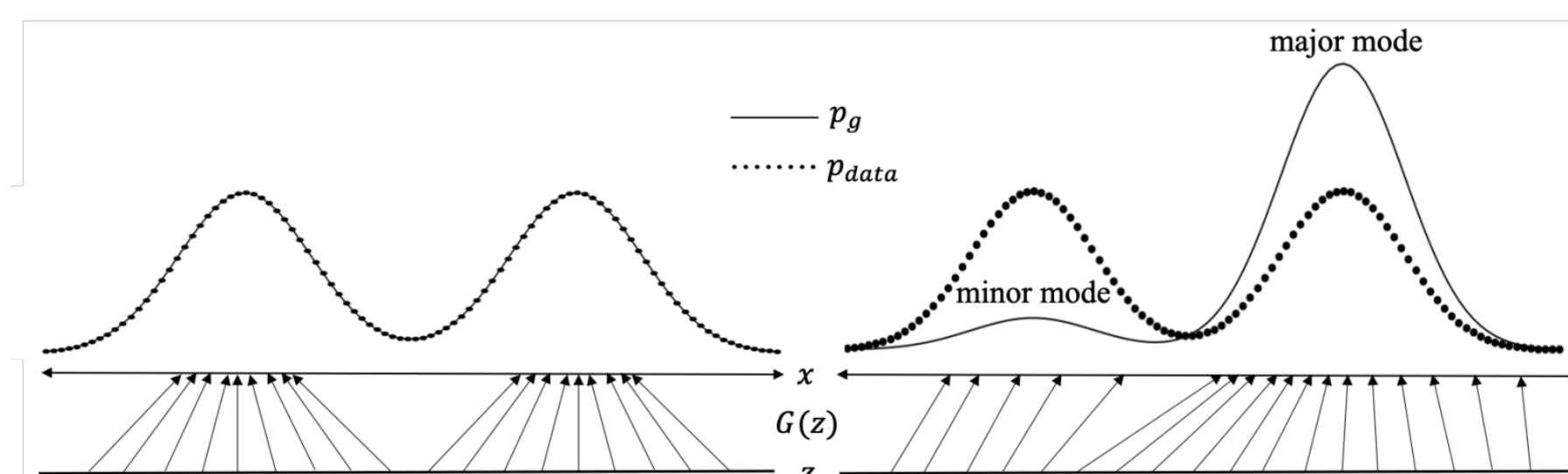
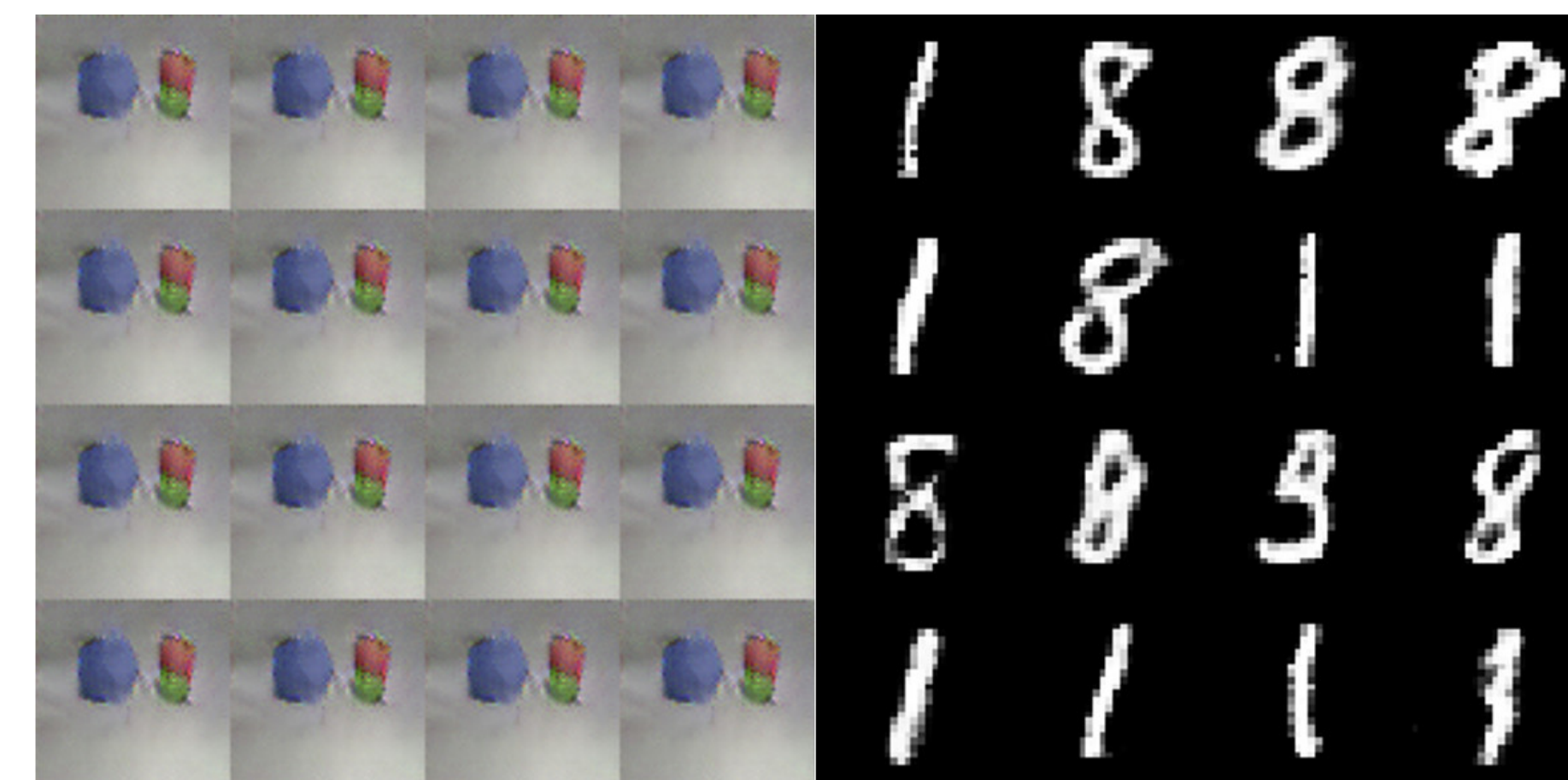


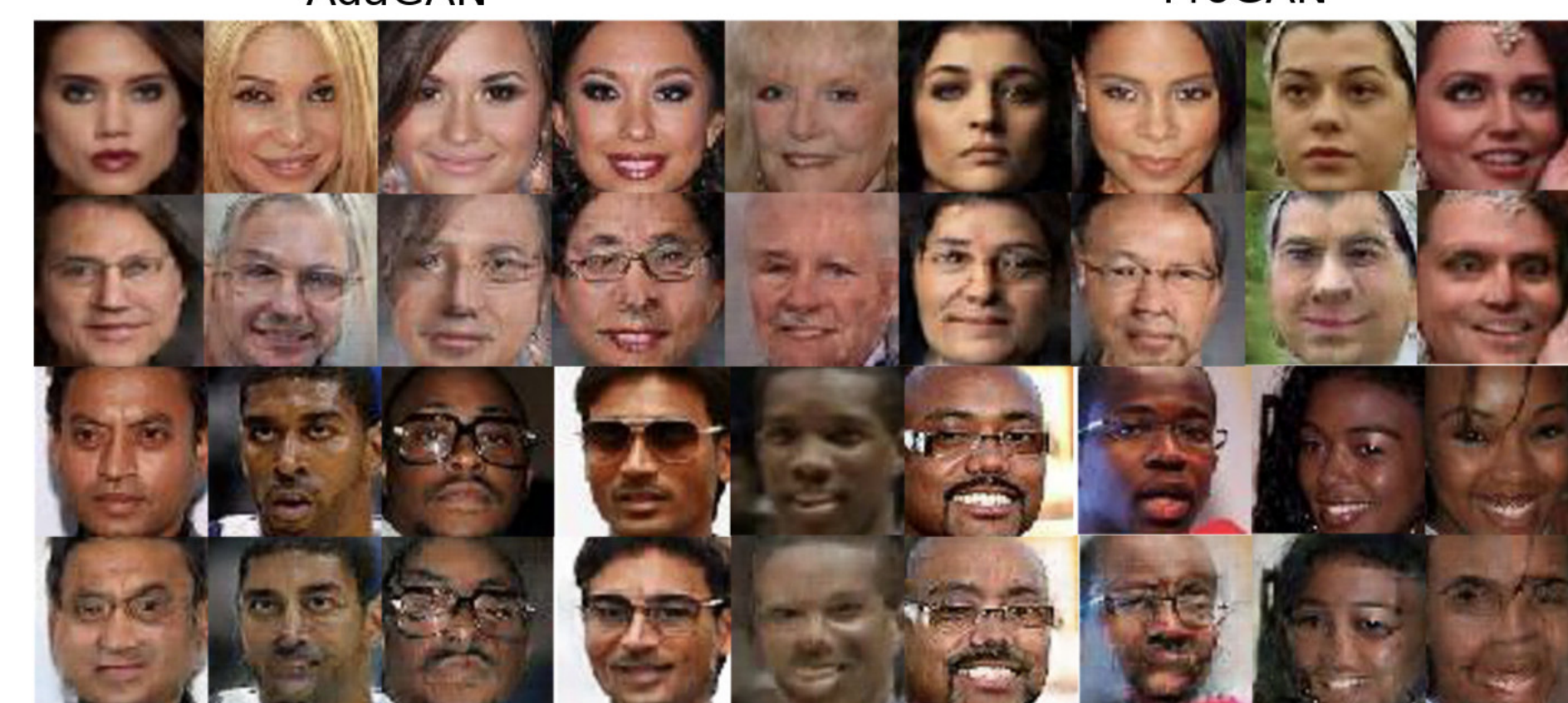
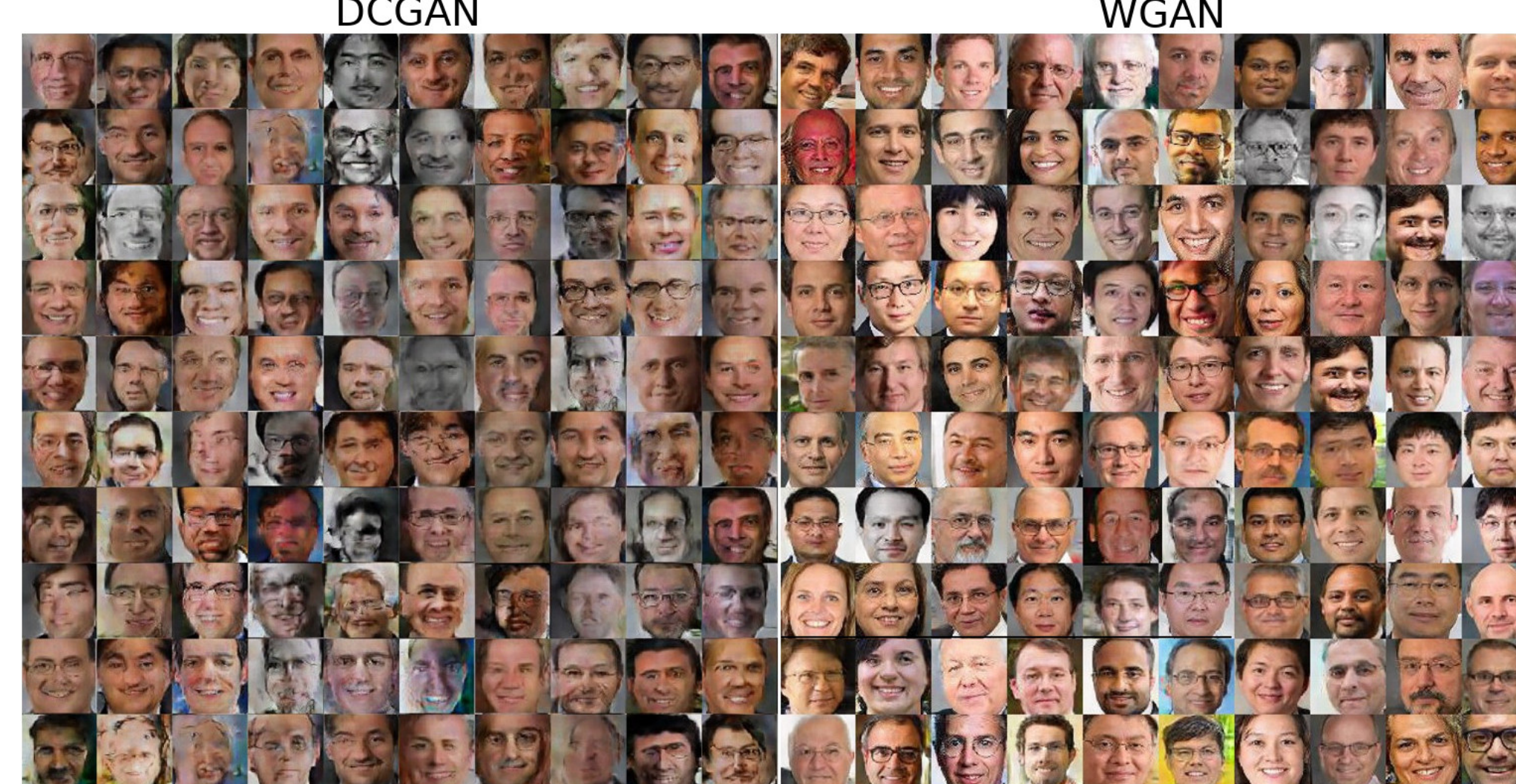
Figure 3: $G(z)$ trained on a uniform, bimodal p_{data} without and with mode collapse. Figure inspired by Goodfellow et al. 2014, Figure 1



Complete Partial

Evaluation

- Gathering and preprocessing of engineering professor headshots dataset from 47 U.S. universities
- Training and generation of new headshot distributions from 4 unconditional and 1 conditional GANs



CycleGAN

- Human annotation tasks on 50 images from each dataset
- Microsoft's Face API: gender recognition

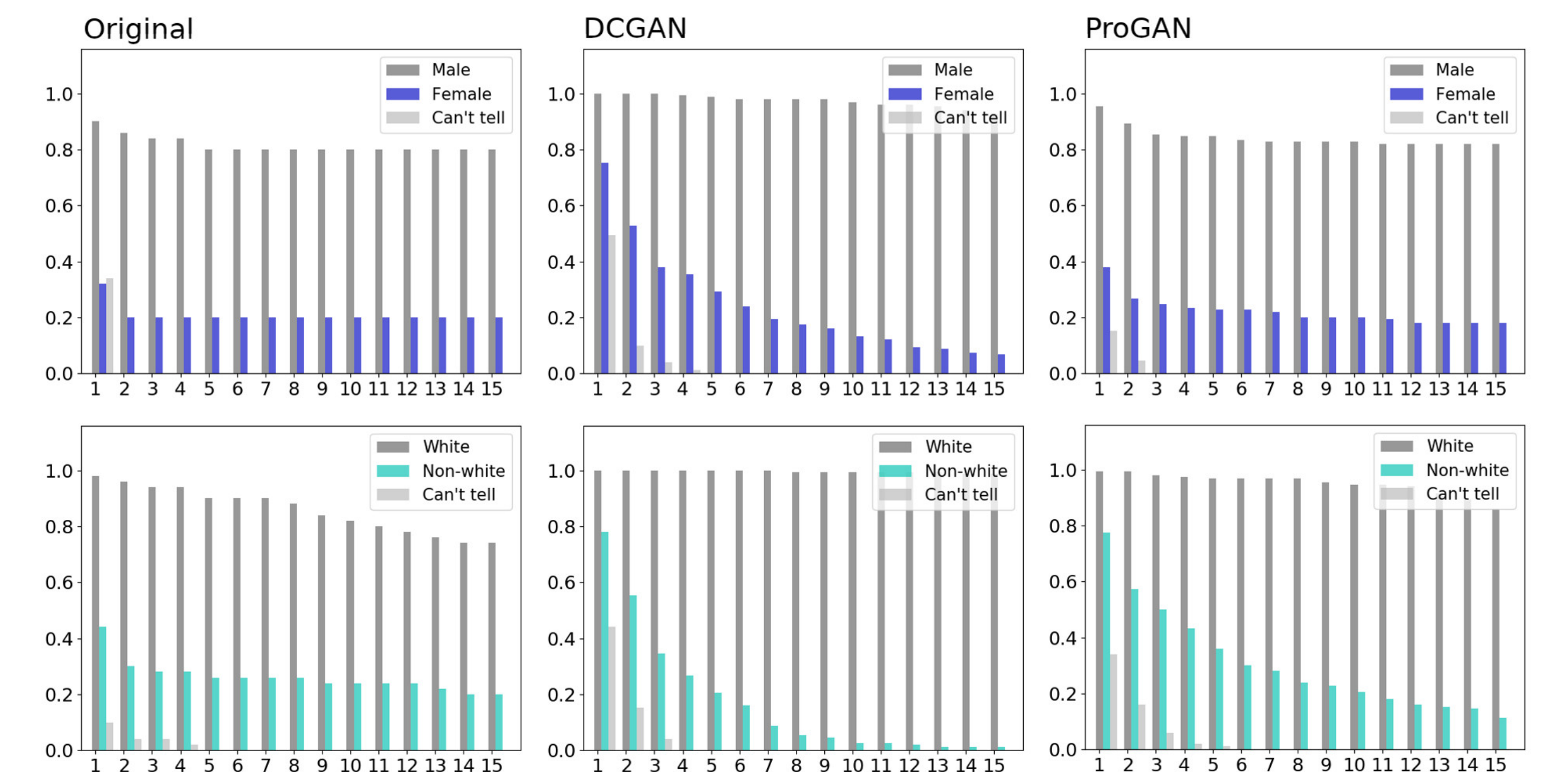
Results

Human annotation

Feminine | Masculine features



Non-white | White color



MS Face API

Feminine | Masculine | Can't tell features

