Imperfect ImaGANation:
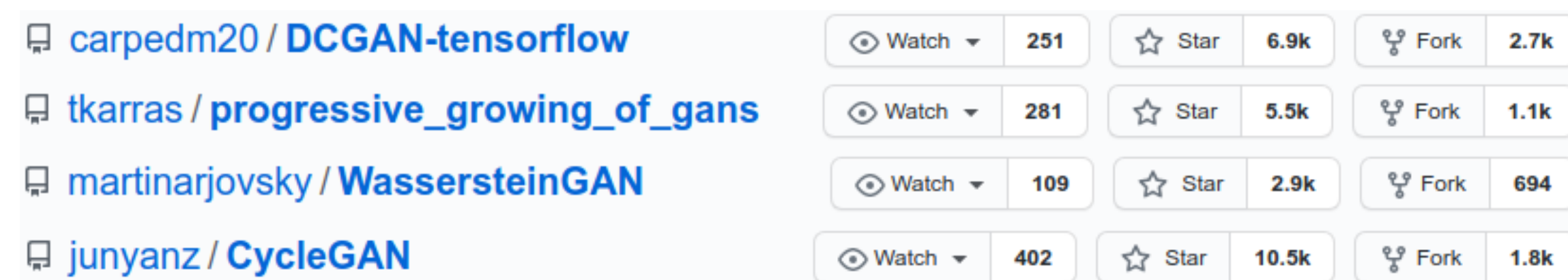# Implications of GANs Exacerbating Biases on Facial Data

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.

Complete   Partial
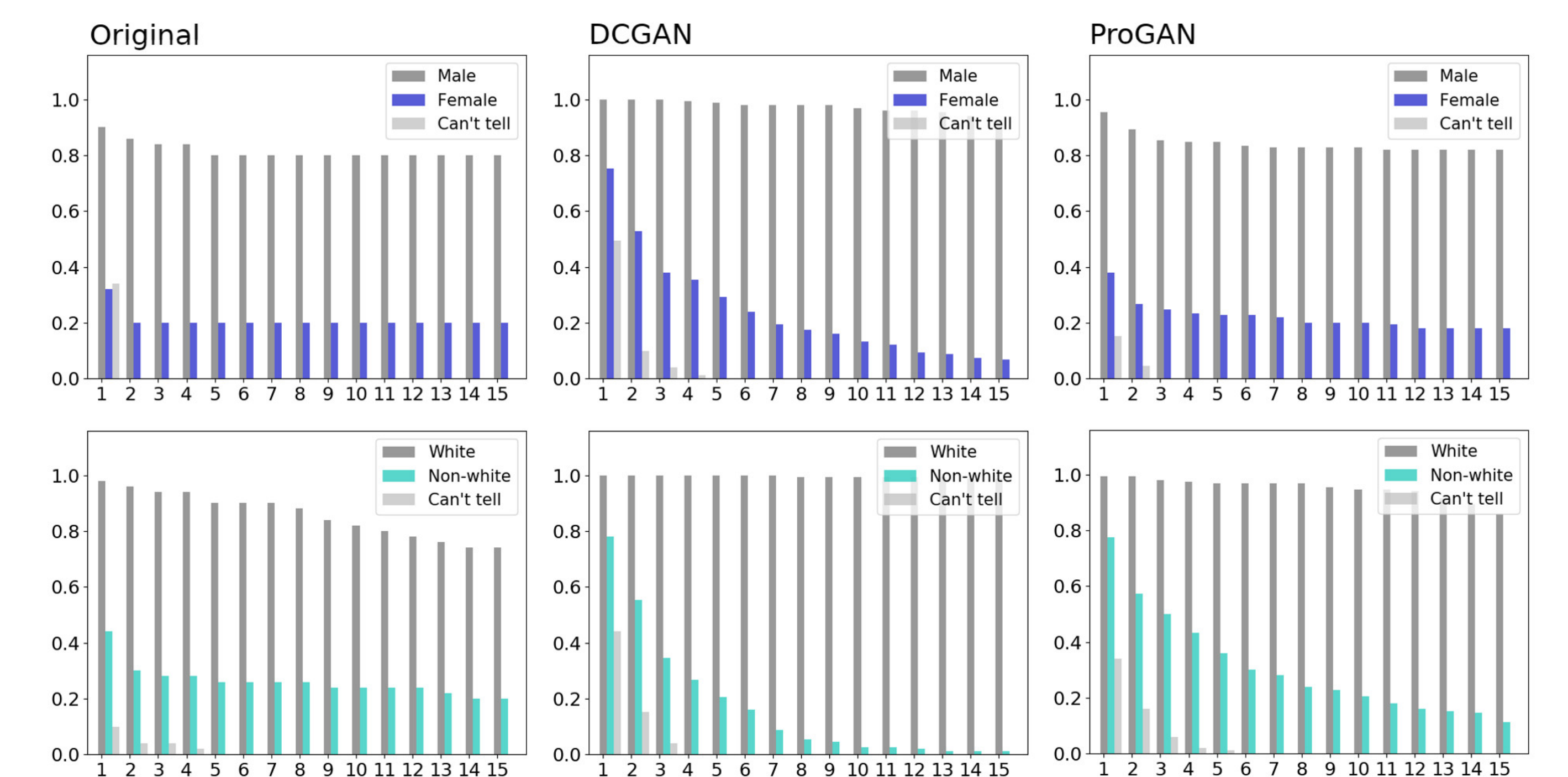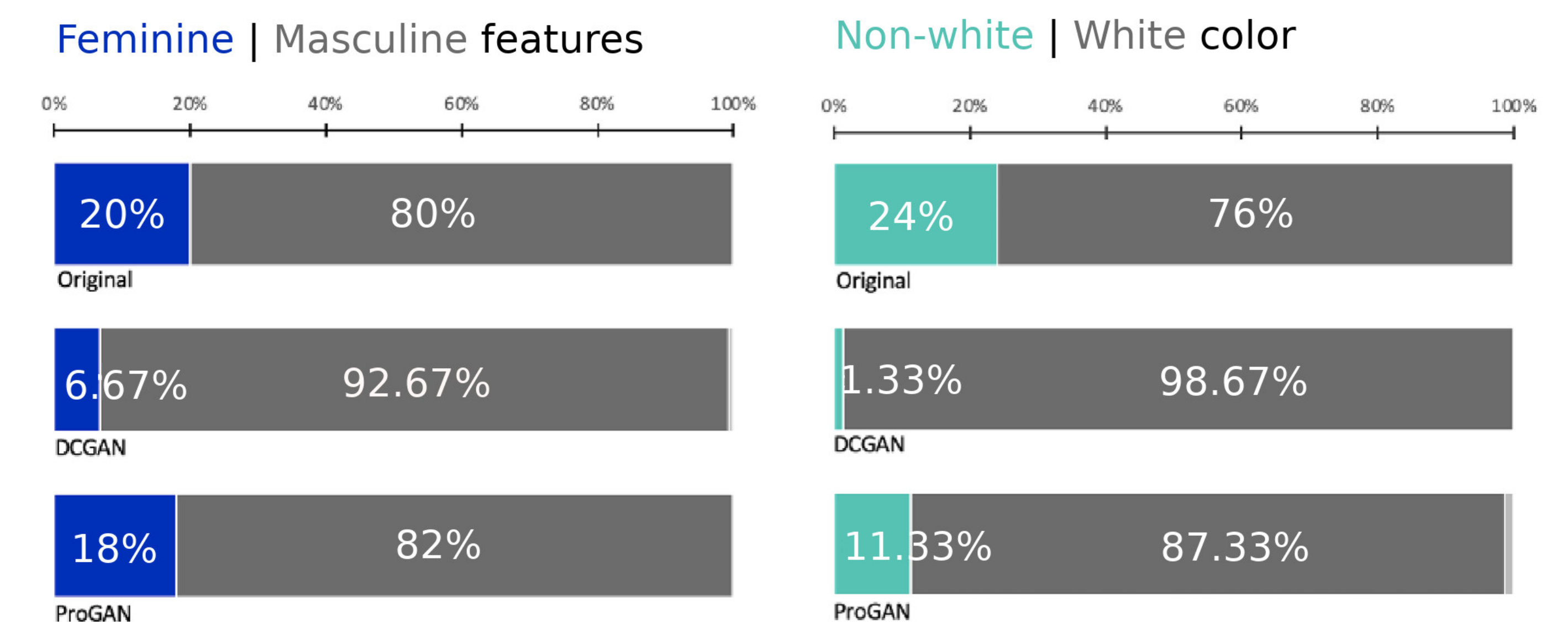

Arizona State University
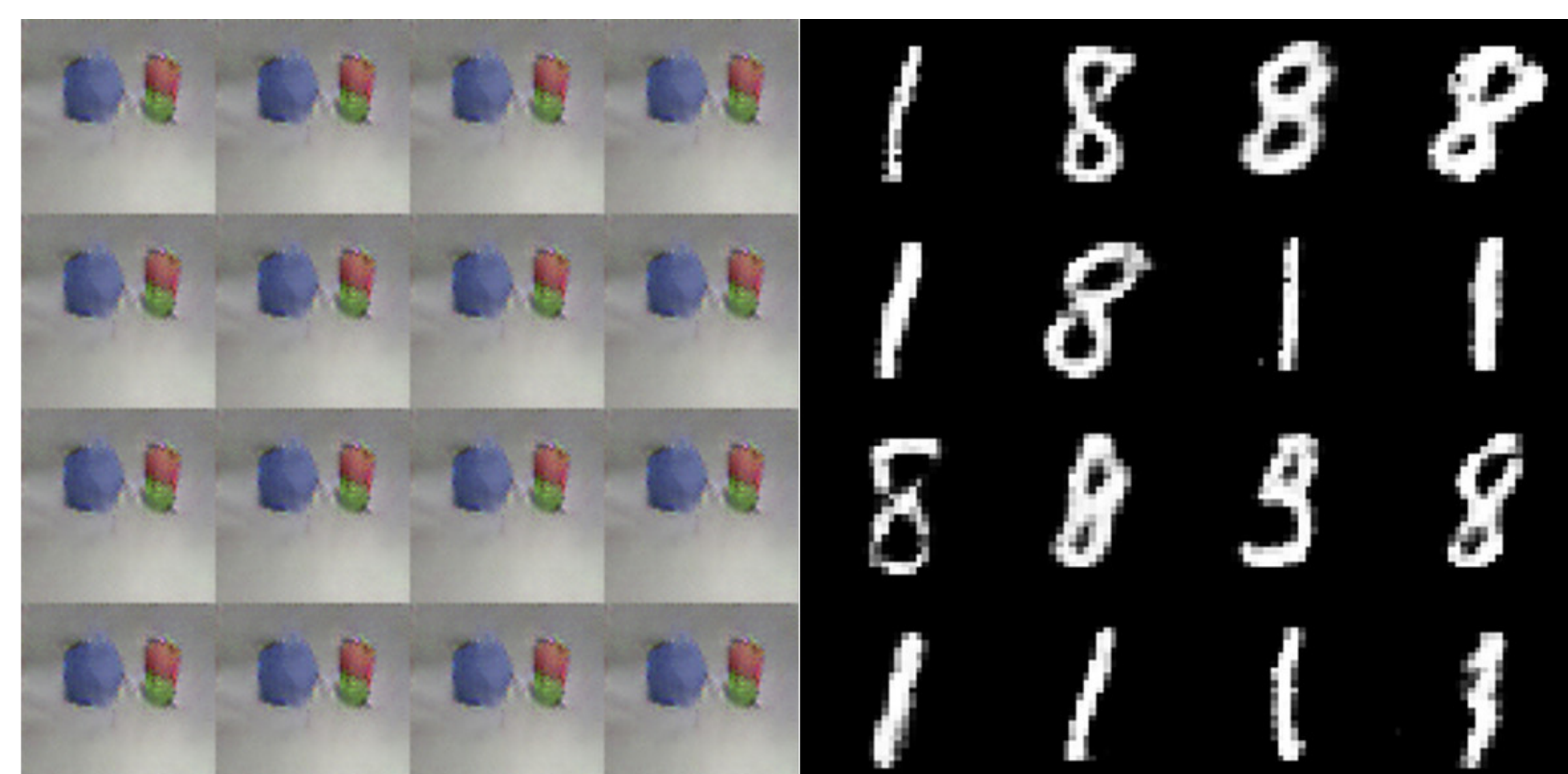
Evaluation

● 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

DCGAN   WGAN

AdaGAN   ProGAN

● Human annotation tasks on 50 images:
  ◇ **T1a** gender on random professor images
  ◇ **T1b** gender on GAN-generated images
  ◇ **T2a** race on random professor images
  ◇ **T2b** race on GAN-generated images

● Microsoft's Face API: gender recognition

✈ njain30@asu.edu | aolmoher@asu.edu

Results

Human annotation

Feminine | Masculine features   Non-white | White color

Original: 20% / 80%   Original: 24% / 76%
DCGAN: 6.67% / 92.67%   DCGAN: 1.33% / 98.67%
ProGAN: 18% / 82%   ProGAN: 11.33% / 87.33%

Original   DCGAN   ProGAN

MS Face API

Feminine | Masculine | Can't tell features

	Feminine	Masculine
Original	16.5%	80.4%
DCGAN	11.8%	87.1%
AdaGAN	3.9%	84.3%
ProGAN	14.8%	82.9%