(Optional Notebook) Score-based Generative Modeling

*Please note that this is an optional notebook meant to introduce more advanced concepts. If you’re up for a challenge, take a look and don’t worry if you can’t follow everything. There is no code to implement—only some cool code for you to learn and run!*

***Click*** [***this link***](https://colab.research.google.com/github/https-deeplearning-ai/GANs-Public/blob/master/C2W2_(Optional_Notebook)_Score_Based_Generative_Modeling.ipynb) ***to access the Colab notebook!***

This is a hitchhiker's guide to score-based generative models, a family of approaches based on [estimating gradients of the data distribution](https://arxiv.org/abs/1907.05600). They have obtained high-quality samples comparable to GANs (like below, figure from [this paper](https://arxiv.org/abs/2006.09011)) without requiring adversarial training, and are considered by some to be [the new contender to GANs](https://ajolicoeur.wordpress.com/the-new-contender-to-gans-score-matching-with-langevin-sampling/).

Machine Bias

Before going into the discussion on bias in machine learning, please read this case study to gain an understanding of the impact these biases can have on real lives:

Machine Bias (Angwin, Larson, Mattu, and Kirchner, 2016): <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

# Fairness Definitions

To understand some of the existing definitions of fairness and their relationships, please read the following paper and view the Google glossary entry for fairness:

1. Fairness Definitions Explained (Verma and Rubin, 2018): <https://fairware.cs.umass.edu/papers/Verma.pdf>
2. Machine Learning Glossary: Fairness (2020): <https://developers.google.com/machine-learning/glossary/fairness>

# A Survey on Bias and Fairness in Machine Learning

To understand the complex nature of bias and fairness, please read the following paper describing ways they exist in artificial intelligence and machine learning:

A Survey on Bias and Fairness in Machine Learning (Mehrabi, Morstatter, Saxena, Lerman, and Galstyan, 2019): <https://arxiv.org/abs/1908.09635>

Finding Bias

Now that you've seen how complex fairness is, how do you find bias in existing material (models, datasets, frameworks, etc.) and how can you prevent it? These two readings offer some insight into how bias was detected and some avenues where it may have been introduced.

1. Does Object Recognition Work for Everyone? (DeVries, Misra, Wang, and van der Maaten, 2019): <https://arxiv.org/abs/1906.02659>
2. What a machine learning tool that turns Obama white can (and can't) tell us about AI bias (Vincent, 2020): <https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias>

Works Cited

All of the resources cited in Course 2 Week 2, in one place. You are encouraged to explore these papers/sites if they interest you, especially because this is an important topic to understand. They are listed in the order they appear in the lessons.

From the videos:

* Hyperspherical Variational Auto-Encoders (Davidson, Falorsi, De Cao, Kipf, and Tomczak, 2018): <https://arxiv.org/abs/1804.00891>
* Generating Diverse High-Fidelity Images with VQ-VAE-2 (Razavi, van den Oord, and Vinyals, 2019): <https://arxiv.org/abs/1906.00446>
* Conditional Image Generation with PixelCNN Decoders (van den Oord et al., 2016): <https://arxiv.org/abs/1606.05328>
* Glow: Better Reversible Generative Models (Dhariwal and Kingma, 2018): <https://openai.com/blog/glow/>
* Machine Bias (Angwin, Larson, Mattu, and Kirchner, 2016): <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
* Fairness Definitions Explained (Verma and Rubin, 2018): <https://fairware.cs.umass.edu/papers/Verma.pdf>
* Does Object Recognition Work for Everyone? (DeVries, Misra, Wang, and van der Maaten, 2019): <https://arxiv.org/abs/1906.02659>
* PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models (Menon, Damian, Hu, Ravi, and Rudin, 2020): <https://arxiv.org/abs/2003.03808>
* What a machine learning tool that turns Obama white can (and can't) tell us about AI bias (Vincent, 2020): <https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias>

From the notebook:

* Mitigating Unwanted Biases with Adversarial Learning (Zhang, Lemoine, and Mitchell, 2018): <https://m-mitchell.com/papers/Adversarial_Bias_Mitigation.pdf>
* Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision at CVPR 2020 (Gebru and Denton, 2020): <https://sites.google.com/view/fatecv-tutorial/schedule?authuser=0>
* Machine Learning Glossary: Fairness (2020): <https://developers.google.com/machine-learning/glossary/fairness>
* CelebFaces Attributes Dataset (CelebA): <http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>