Comparing the interpretability of different GAN architectures

Henry Tu
University of Toronto

Seel Patel
University of Toronto

1   Abstract

A Generative Adversarial Network (GAN) uses an adversarial process between two models which are simultaneously trained to estimate a generative model.[1] There are many variants of GAN architectures, such as COCO-GAN[2] and StyleGAN[3], which both have the ability to generate synthetic images which mimic real images.

We are exploring methods of comparing the quantitative performance of these architectures with their interpretability. Our quantitative measures include C2ST[4], image quality measures[4], Maximum Mean Discrepancy (MMD)[4], etc.

In order to analyze these networks qualitatively, we will use methods such as SmoothGrad[5], Latent Space Exploration[6] and nearest neighbour tests[4] to decipher what the networks have learned.

By combining these two forms of analysis, we hope to gain insight into the relationship between performance metrics and the generated output of the models.

2   Introduction

StyleGAN and COCO-GAN generate images using different techniques to accomplish two different goals: StyleGAN is designed such that high level features of the output image can be finely tuned and adjusted (e.g. lighting, hair colour, etc.)[3]. On the other hand, COCO-GAN generates each part of the image separately before stitching it together in order to simulate the human perception of vision[2]. As a result, we expected each model to generate images with different qualities that relate to their designed tasks.

Although we tried to analyze the models using all the metrics listed in the Abstract, there were a few challenges involved with getting the desired results.

3   Quantitative Analysis

3.1   C2ST

asd

3.2   Image Quality

asd

3.3   MMD

asd

4 Qualitative Analysis

4.1 SmoothGrad

asd

4.2 Latent Space Exploration

asd

4.3 Nearest Neighbours

asd

References

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. (2014) *Generative Adversarial Networks*

[2] Chieh Hubert Lin, Chia-Che Chang, Yu-Sheng Chen, Da-Cheng Juan, Wei Wei, Hwann-Tzong Chen. (2019) *COCO-GAN: Generation by Parts via Conditional Coordinating*

[3] Tero Karras, Samuli Laine, Timo Aila. (2018) *A Style-Based Generator Architecture for Generative Adversarial Networks*

[4] Brownlee, Jason. (2019) *How to Evaluate Generative Adversarial Networks.*

[5] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viegas, Martin Wattenberg. (2017) *SmoothGrad: removing noise by adding noise*

[6] Tom White. (2016) *Sampling Generative Networks*