GAN Papers from Nvidia

April 30, 2019

In 2017 and 2018, researchers at Nvidia produced two significant papers on generative adversarial networks (GANs) [2, 3]. These papers results in automatic generation of high quality and high resolution of images of human faces. (See http://thispersondoesnotexist.com, for example.) This note is written as I read those papers.

1 StyleGAN Paper

This is the "A Style-Based Generator Arcthitecture for Generative Adversarial Network" paper [3], which was published to CVPR 2019.

1.1 High-Level Description

- The paper describes a new architecture for GAN's generator, which is inspired by the networks for style transfer [1].

- Typical GANs start with a latent code, which is then transformed directly by the network into an output.

- However, the paper's architecture starts with a learned constant input which is adjusted to the style given by the latent code as it progresses through the network.

- The constant input starts at a low resolution. It is then adjusted for style and then upsampled. This is repeated for a number of times until the desired resolution is achieved. The upsampling is inspired by the previous GAN paper [2].

- The paper claims that the new generator can separate high-level attributes (pose, identity, etc.) from low-level ones (freckles, hair, etc.).

- One insight of the paper is the benefit of distribution of latent codes from the real distribution. In previous GANs, the distribution of the latent codes must follow that of the training data. Using the latent codes to adjust styles leads the two distributions to become less entangled.

    - The paper introduces two metrics—perceptual path length and linear separability—for measuring disentanglement between two distributions.
    - Using the new metrics, the paper shows that their generators yields a more linear, less entanbled representatin of different factors of variation.

- The paper also provides a new dataset of human faces called the Flickr-Faces-HQ (FFHQ) dataset.

1.2 The Generator

- For a given latent code $\mathbf{z} \in \mathcal{Z}$, the paper uses a non-linear mapping network $f : \mathcal{Z} \to \mathcal{W}$ to produce $\mathbf{w} \in \mathcal{W}$.

  - Both $\mathbf{z}$ and $\mathbf{w}$ are vectors of size 512.
  - The mapping f is implemented with an 8-layered multi-layer perceptron.

- Learned affine transformations then transform $\mathbf{w}$ to styles $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_b)$ that contains adaptive instance normalization (AdaIN) layers [1]:

$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i}.$$

References

[1] HUANG, X., AND BELONGIE, S. J. Arbitrary style transfer in real-time with adaptive instance normalization. *CoRR abs/1703.06868* (2017).

[2] KARRAS, T., AILA, T., LAINE, S., AND LEHTINEN, J. Progressive growing of gans for improved quality, stability, and variation. *CoRR abs/1710.10196* (2017).

[3] KARRAS, T., LAINE, S., AND AILA, T. A style-based generator architecture for generative adversarial networks. *CoRR abs/1812.04948* (2018).