Proposal of Implementation and Improvement of StyleGAN

Fei Zheng fz2277 [*1] **Chirong Zhang cz2533** [*1] **Xiaoxi Zhao xz2740** [*1]

1. Previous work and references

The Style-Based Generator Architecture for Cenerative Adversarial Networks (GAN)(Karras et al., 2019) has proposed a new generator architecture for GAN using style transfer techniques(Gatys et al., 2016). This new architecture disentangles the latent factors of variation, which is one of the main limitations of ProGAN(Karras et al., 2018). Here are the main breakthroughs:

- A style-based generator with unsupervised separation of high-level attributes
- Scale-specific control of synthesis
- Generator starts from trainable constants and embeds the input latent code into an intermediate latent space which better disentangles the features
- Two new metrics to quantify the disentanglement: perceptual path length and linear separability
- Present a new dataset of human faces with wider variation and higher quality

2. New Problem Proposal: Style Mixing

The Style GAN(Karras et al., 2019) employ mixing regularization to localize the styles and operate style mixing with two latent codes Z_1, z_2.

Instead of just mixing two styles, we want to run 3 latent codes through the mapping network to mix multiple styles in our generator. We will have the corresponding w_1, w_2, w_3 applied between different crossover points.

Initially, we choose three number of latent codes because coarse spatial resolutions $(4^2 - 8^2)$ brings high-level aspects such as pose, face shape, as middle resolutions $(16^2 - 32^2)$ brings hair style, eye open/closed and fine styles $(64^2 - 1024^2)$ brings color scheme and other microstructure.

[*]Equal contribution [1]Department of Statistics, Columbia University, New York, USA. Correspondence to: Iddo Drori <id2305@columbia.edu>.

Final Project of COMS4995 Deep Learning, Columbia University, New York. Copyright 2019 by the author(s).

We will first apply this style mixing to human faces and then apply it to animal datasets as well, to answer what will my cats look like if they become dogs and to generate photos of potential new breeds/species.

Then if time permits, we will further try to raise the number of styles and apply this to other datasets as well. (food, art) We will also try to extract one single aspect (e.g. nose) and only stylize that specific aspect. (e.g. change one's nose into a Joker's nose style)

2.1. Evaluation Criteria

We will use Frechet inception distance (FID)(Heusel et al., 2018)for various generator designs (lower is better) ans also for each generated photos, we will calculate the style loss (Gatys et al., 2016).

2.2. Dataset

We will first apply our style mixing on a dataset of human faces (Flickr-Faces-HQ, FFHQ) [1], then apply on the animals dataset [2] [3]

References

Gatys, L. A., Ecker, A. S., and M.Bethge (eds.). *Image Style Transfer Using Convolutional Neural Networks*. IEEE, 2016.

Heusel, M., Ramsauer, H., T. Unterthiner, B. N., and Hochreiter, S. (eds.). *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. 2018.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (eds.). *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. 2018.

Karras, T., Laine, S., and Aila, T. (eds.). *A Style-Based Generator Architecture for Generative Adversarial Networks*. 2019.

[1]https://github.com/NVlabs/stylegan

[2]http://www.robots.ox.ac.uk/ vgg/data/pets/

[3]https://www.kaggle.com/alessiocorrado99/animals10