Proposal: Text-to-image Implementation of Conditional Style GAN

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

1. Previous work and references

1.1. Conditional GAN

Conditional GAN(Isola et al., 2018) has investigated a general-purpose solution to image-to-image translation problems. It introduces conditional information both in generator and discriminator, which makes the generation process more supervised.

1.2. Style GAN

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: Text-to-image Translation

The pre-trained model BERT(Devlin et al., 2018) made it a great breakthrough in the domain of word embedding. Conditional GAN(Isola et al., 2018) proposed a new solution to translating from images to images, which might be applicable in translating from other than images. Style GAN(Karras et al., 2019) investigated a new architecture to generate high-quality images.

We try to combine advantages of these networks in our network and apply it to translate texts to corresponding images.

Initially, we will use pre-trained BERT(Devlin et al., 2018) to transform texts to embedding vectors. After concatenating it with random noise from normal distribution, we will run it through the mapping network and the generator, just the same as in style GAN(Karras et al., 2019). Then we put the tuple of text and fake image and the corresponding tuple of text and real image into discriminator, just like in conditional GAN.

Since there are several alternative loss functions in these researches, this is a main field we would explore. We might consider inception distance, L1 distance or L2 distance in generating process. We will first train our model on the flower dataset. Our target is to generate a corresponding flower image given any description of the flower.

Then if time permits, we will extend our model on a dataset including more common object to see whether our model can generalize to much more broad cases.

2.1. Evaluation Criteria

We will use Frechet inception distance (FID)(Heusel et al., 2018)(lower is better) to calculate how far the generated images are away from the real images

2.2. Dataset

We will apply our text-to-image model on a dataset of flowers with 102 categories and 5 captions for each image [1]. Then we may extend our model to COCO dataset [2] including common objects.

[*]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.

[1]http://www.robots.ox.ac.uk/ vgg/data/flowers/102/
[2]http://cocodataset.org/

References

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (eds.). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018.

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.

Isola, P., Zhu, J., Zhou, T., and Efros, A. A. (eds.). *Image-to-Image Translation with Conditional Adversarial Networks*. 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.