CIS 5200 Fall 2022, Project Proposal

Latent Space Structure between VAEs on Different Data

Team Members:

- Yimeng Zeng; Email: yimengz@seas.upenn.edu
- Huanming Song; Email: noeland@seas.upenn.edu
- Haosong Huang; Email: hhuang2@seas.upenn.edu

home pod: Joyful Jellyfish

---

1 Motivation

Variational autoencoders (VAEs) [3] are known for their ability to synthesize data from Gaussian noise. VAEs have the nice property that it tries to organize the latent space such that the latent codes generated are Gaussian distributed. Since we are assuming that the VAEs are learning some of the underlying distributions of the dataset, it is natural to ask, how well can the learned mapping from data to latent space transfers across tasks. For example, if we train a VAE on a human faces dataset, pass an image through the encoder, and then feed the generated latent code to a decoder trained on an anime faces dataset, will it decode to an actual anime drawing similar to the human face, or will the underlying distribution learned be very different. Motivated by this question, we choose to investigate how are the latent space of VAEs structured across similar and different tasks.

2 Related Work

There are numerous works done on image-to-image translation, some of them utilizing VAEs and GANs at the same time [2], some relying on energy-based models [5]. There also has been various improvements made to VAEs themselves, for example, $\beta-$VAEs [1] which introduces a hyperparameter β which tunes the degree of disentanglement learned by the encoder. InfoVAE [4], which utilizes the latent space more efficiently regardless of the flexibility of the decoding distribution.

3 Problem Formulation

In this project, we plan to investigate if we can successfully mix and match encoder-decoder pairs trained on similar distributions, and have them produce reasonable outputs. If we find that even if the two distributions are very similar, such as the anime faces[1] and human faces dataset[2], we still have difficulties producing meaningful output, we will then investigate if there are ways in which we can fine-tune the decoder/encoder such that they can learn to decode to different data distribution. If the experiments turn out to work very well, we can then investigate if it is possible to have the VAE decode the latent code of one distribution into another completely unrelated distribution, for example, MNIST and CIFAR-10.

[1] https://www.kaggle.com/datasets/splcher/animefacedataset
[2] https://www.kaggle.com/datasets/ashwingupta3012/human-faces

4 Methods

We will first train several baseline VAEs in the different datasets we have collected, and directly mix and match the encoders together to see how well they can map the latent codes of one distribution to another distribution. Then we will investigate if there are ways we can fine-tune the encoder/decoder with a small number of inputs from the output distribution to allow better-quality image generation. We will also train a different set of VAEs to see if using different restrictions on the latent space will lead to better image generation.

5 Evaluation

The quality of image generation will be mainly human evaluation since there are no exact correctness guarantees. We will look at for example if a man maps to a male anime character or finer details such as whether a picture of a man with sunglasses gets mapped correctly to an anime drawing of a man with sunglasses.

When investigating if our VAE can successfully map MNIST into CIFAR-10, we might first want to set a $1-1$ mapping between classes, and then train a classifier to see how many of the input pictures gets mapped to their corresponding class correctly.

6 Project Plan

Week 11: 11/14

- Download and organize datasets
- Train vanilla VAEs on our datasets
- Have preliminary results on how mixing VAEs together works

Week 12: 11/21

- Train variants of VAEs and evaluate their performance
- Fine tune encoder/decoder for better image generation
- Collect more data if needed

Week 13: 11/28

- Investigate whether VAEs can map from one distribution to another
- Visualizing the latent space of different data distributions and see if there is anything interesting
- Finish writing project report

Week 14: 12/5

- Final project presentation

References

[1] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

[2] Hadi Kazemi, Sobhan Soleymani, Fariborz Taherkhani, Seyed Iranmanesh, and Nasser Nasrabadi. Unsupervised image-to-image translation using domain-specific variational information bound. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[3] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

[4] Shengjia Zhao, Jiaming Song, and Stefano Ermon. Infovae: Information maximizing variational autoencoders. *CoRR*, abs/1706.02262, 2017.

[5] Yang Zhao and Changyou Chen. Unpaired image-to-image translation via latent energy transport. *CoRR*, abs/2012.00649, 2020.