Image Editing with Invertible Conditional GANs

Boyao Sun, Wensheng Li, Xinyu Xu {boysun, we077, xinyu17}@stanford.edu

Motivation

- Generative Adversarial Networks (GANs) has resulted in promising outcomes in image generation in the last few years.
- Invertible Conditional GANs is able to edit the image by reconstructing images with selected attributes
- The whole system consists of a cGAN, a latent representation encoder z, and a conditional information encoder y
- The model works well on digit change, but it does not work well on human face yet the result of the project reflects the feasibility of IcGANs

Data & Features

- The datasets consist of MNIST[1] and CelebA[2]
- MNIST is a greyscale digit dataset with 70,000 images; CelebA is a colored human face dataset with 2 millions

- Input size and output size for MNIST images are 28 x 28; Input size for CelebA images are 178 x 218; output size for CelebA images are 64 x 64
- Images from both datasets are normalized in data preprocessing, and images from CelebA are cropped before resized
- The labels for MNIST images are one-hot vectors, and the labels for CelebA images are vectors with 40 attributes

Models

- The model is built based on DcGAN[4] and IcGAN[3] and is modified to get better loss/visualization (See green parts in Figure 2)
- One fully connected is added in Generator; channel numbers at each layer in Discriminator are increased

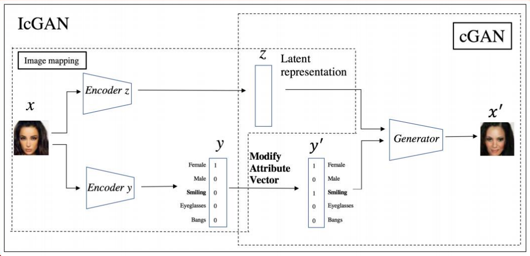

Figure 1: Scheme of a trained IcGAN

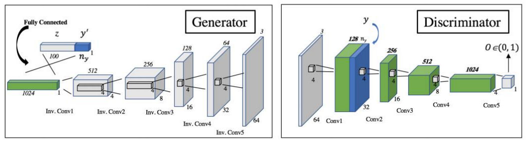

Figure 2: Architecture of the Generator and Discriminator

- Loss Functions are the same as IcGAN; equation (1) is the loss function for the Discriminator and Generator; equation (2) is that for encoder z; equation (3) is that for encoder y

$$min_g max_d v(\theta_g, \theta_d) = \mathbb{E}_{x,y\sim p_{data}}[log D(x,y)] + \mathbb{E}_{z\sim p_z, y'\sim p_y}[log(1 - D(G(z,y'),y'))] \quad (1)$$

$$L_{ez} = \mathbb{E}_{z\sim p_z, y'\sim p_y}\|z - E_z(G(z,y'))\|_2^2 \quad (2)$$

$$L_{ey} = \mathbb{E}_{x,y\sim p_{data}}\|y - E_y(x)\|_2^2 \quad (3)$$

Results

- The reconstructed results for two datasets

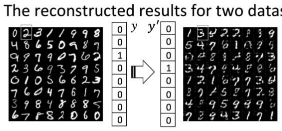

Figure 3a: image for MNIST

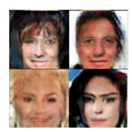

Figure 3b:image for CelebA

- The loss curves for the Generator and Discriminator

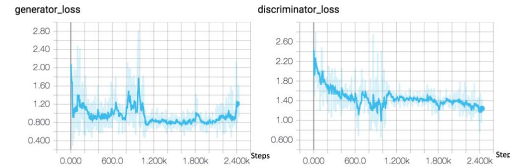

Figure 4: Loss curves for G and D for CelebA

- The 98% of dataset is used for training; 2% of dataset is used for testing

Discussion and Future Work

- We have re-implemented IcGAN on MNIST, which indicates it has successfully modified ¾ digit images as expected. These results illustrate the capacity of IcGAN in image editing.
- Our modified IcGAN has reconstructed some face images.
- As observed in Figure 4, Discriminator loss decreases during training while Generator loss does not have significant reduction.
- This loss meets our expectation because in minimax game, Discriminator becomes powerful and hence Generator loss cannot decrease.
- However, the reconstructed face images are unnatural to us. In addition, our implementation of IcGAN on CelebA does not yield expected results. We believe that the bad performance of IcGAN is because our GAN is unable to output expected images.
- In the future, we would like to put more efforts on training GAN, such as hyperparameters tuning and modifying architectures.
- If GAN is well trained on CelebA and we are given more time, we will implement IcGAN on CelebA dataset and improve its quality.

References

[1] Y. LeCun and C. Cortes. MNIST handwritten digit database, 2010.

[2] Z. Liu, P. Luo, X. Wang, and X. Tang, Deep learning face attributes in the wild. *In Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[3] G. Perarnau, J. van de Weijer, B. Raducanu, and J.M. Alvarez. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*, 2016.

[4] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv: 1511.06434*, 2015.